

Finding a future for GM crops

Some parties to debates over the safety of genetically modified crops don't want to hear the results of objective research. But that isn't an excuse for not doing it.

Of all of the tactics employed by the opponents of genetically modified (GM) crops in Europe, one of the most effective has been to publicize the connections between private companies selling agricultural biotechnology and the scientists assessing the safety of the crops.

For those unfamiliar with how research is funded, this information can create the impression of a conflict of interest. Many of the scientists involved serve as advisers to the companies. Others carry out research funded by the firms. These links need not undermine the integrity of anybody's research, but they have nonetheless undermined public confidence in that integrity.

Governments seem unperturbed by the damage caused by this impression. In Britain, for example, an industry-funded body, the Supply Chain Initiative on Modified Agricultural Crops, has helped to coordinate farm-scale trials of the crops' impact on biodiversity. Industry has a similar role in trials elsewhere in Europe. In the United States, where the crops are already grown in large quantities, federal agencies have been happy to let the industry itself carry out most of the research into the safety and environmental impact of GM crops.

There's nothing intrinsically wrong with involving industry in such pre-regulatory research. Those doing the research gain from the considerable expertise of the corporations that have developed the transgenic crop strains, and the taxpayer benefits from having industry foot the bill. But the arrangement can undoubtedly provide ammunition for those who later attempt to discredit the studies. And if industry's role in the work is predominant, this criticism may have a point.

This September, for example, the Agriculture and Environment Biotechnology Commission (AEBEC), an independent body advising the British government, reported that the range of indicators of biodiversity that are being used in the UK farm-scale trials is too narrow to assess the crops' impact properly.

The AEBEC didn't attack the impartiality of the scientists involved

or the quality of their work. There was no reason to do so: if the work is carried out at independent research institutes and is properly peer reviewed, there is no reason why industry's funding should bias its results. The problem lies not in the individual projects, but rather in the range of questions they address. If the agricultural biotechnology industry sets the questions — as appears to have happened in this case — then the research is unlikely to fulfil the public's interest.

The British government needn't look too far to fix this problem. Its own Food Standards Agency (FSA) has a similar remit to the AEBEC's in terms of providing strategic advice. But it also has a modest research budget of its own, £21 million (US\$30 million) a year, so that it can commission the research it deems necessary. The government should consider expanding the AEBEC's remit along similar lines, initially, perhaps, giving it a mandate to draw up a framework for the GM-crop-related research that is required in the United Kingdom.

In the United States, regulation of GM crops is divided between three agencies — the Food and Drug Administration, the Environmental Protection Agency and the US Department of Agriculture — depending on the genetic modification to the crop. But all three agencies are severely constrained in their ability to set a research agenda that will protect the public interest. Funding for research on GM technology is too heavily concentrated in the hands of producers to be compatible with that objective.

The wide acceptance of GM crops in the United States, and their likely rejection in Europe regardless of what the studies say, might suggest that reforming research to give better information is an academic exercise. But it is still worth doing. The results of past and future trials will feed into the looming trade dispute over whether European rules on issues such as labelling of crops are protectionist. And new GM crops will have to be assessed as the technology advances. If the research agenda is to help governments meet these challenges, it must be set independently of industrial influence, and seen as such. ■

Double standards

Developing countries deserve as much leeway to deal with their public-health crises as the United States has allowed itself.

The next ministerial meeting of the World Trade Organization in Doha, Qatar, on 9–13 November promises to be a lively affair. One agenda item, in particular, that will demand attention is an expected appeal by developing nations, led by Brazil, for greater leeway in the compulsory licensing of pharmaceuticals in the event of public-health emergencies, under the TRIPS (Trade-Related Intellectual Property Rights) agreement.

Brazil, the advocacy group Doctors Without Borders and many African governments have for some time been pushing for greater use of a provision in TRIPS that allows nations to override patent protection on drugs in "national emergencies". The US government has opposed this flexibility, sometimes threatening to challenge the use of the provision by developing countries facing AIDS crises.

It is salutary, then, to see how America has responded to a national emergency of its own — the recent anthrax attacks — that required

action on drug affordability. Initially, Canada announced that it would license the production of Cipro (ciprofloxacin), a treatment for anthrax, by suppliers other than the drug's patent holder, Bayer. Then Senator Chuck Schumer (Democrat, New York) began to appear on television demanding that the United States do the same.

The Bush administration at first dismissed such action as illegal. But, forced by the persistent senator to acknowledge that this was not the case, it proceeded to extract agreement from Bayer to supply the drug at about one-fifth of its previous price. The health secretary, Tommy Thompson, even boasted that the threat of compulsory licensing had helped to clinch the deal.

No such muscle is available to the smaller African nations, whose populations are being ravaged by AIDS. They should demand as much leeway in interpreting TRIPS as the United States has seen fit to allow itself. ■





Private lives
Surveillance powers
spark arguments
over privacy
p4



Lucky stars
Bungled experiment
sheds light on
planetary formation
p6



No holds baa'd
Sheep under scrutiny
as Britain seeks out
signs of BSE
p7



Fishy friend
Compact puffer-fish
genome boosts gene
counting effort
p8

Crackdown on hazardous agents raises concern for bona fide labs



Killer question: will the new laws hamper labs that are building defences against bioterrorism?

Jonathan Knight, San Francisco

The United States is moving to place tighter controls on the possession and transport of hazardous biological agents. But biologists and the societies that represent them are worried by the extent of the clampdown.

In testimony before a Senate panel this week, representatives of the biologists will seek to soften proposed restrictions that would stop foreign nationals from handling dangerous agents in US laboratories. They argue that the restrictions go too far, and could disrupt research that would help to build defences against biological attack.

Since 1997, US laboratories that ship or receive any of 36 harmful viruses, bacteria, fungi and toxins listed as "select agents" in the 1996 Antiterrorism and Effective Death Penalty Act have been required to register with the Department of Health and Human Services. Among these select agents are anthrax, ebola virus, botulinum toxin and *Yersinia pestis*, the bacterium that causes plague.

Even before the latest security crisis, pressure was growing for far tighter regulation of the handling of such agents. In 1999, for example, the Clinton administration and many Republicans in Congress supported changes that would have strengthened the rules. But legislation to enact the changes

floundered, reportedly in the face of opposition from universities.

The anti-terrorism act that was signed into law by President George W. Bush on 26 October (see page 4) bans foreign nationals from seven countries accused by the State Department of supporting terrorism — Cuba, Iran, Iraq, Libya, North Korea, Sudan and Syria — from handling the agents in the United States.

A second bill, the Bioterrorism Enforcement Act, which has been passed by the House of Representatives but not the Senate, would prevent all non-citizens other than permanent residents from working with the listed agents — although the health secretary would be allowed to make exceptions. The provision worries US biologists because so many of their laboratories employ foreign nationals.

"It's a real concern to me that the rules will be tightened to the point of causing serious restrictions on our research," says John Collier, a microbiologist at Harvard Medical School in Boston, whose papers on the mechanism of anthrax infection were published last week on *Nature's* website.

Janet Shoemaker, director of public and scientific affairs at the American Society for Microbiology (ASM), said that the society had negotiated with Senator Patrick Leahy (Democrat, Vermont) to include "bona fide

Atomic-bomb experts interrogated over Taliban links

David Adam, London

Two retired nuclear-weapons physicists were questioned last week by Pakistani security services over their possible links to the Taliban regime in Afghanistan.

A spokesperson for the Pakistani Ministry of Foreign Affairs confirmed that Sultan Bashiruddin Mahmood, former director of the Pakistan Atomic Energy Commission (PAEC), which built the nation's atomic bomb, and Chaudry Abdul Majeed, a former senior scientist at the PAEC, were "asked some questions" about their connections

with Afghanistan. The duo were reportedly taken into custody in Lahore on 23 October.

"They have not been arrested and they are not under detention," the spokesperson said, adding that they were questioned over their involvement with an Afghan relief agency founded by Mahmood. "It is just an investigation to look into the credentials of various non-governmental organizations who may have been working inside Afghanistan," the spokesman added.

Mahmood was released earlier this week after being cleared by the security agencies,

according to officials quoted in the Pakistani newspaper *Dawn*, which described him as "a staunch supporter" of the Taliban. Some observers have expressed concern about the combination of Mahmood's technical expertise and his political sympathies.

However, Bashir Syed, former president of the Association of Pakistani Scientists and Engineers of North America, said: "I know both of these persons and can tell you there is not an iota of truth that both these respected scientists and friends will do anything to harm the interest of their own country." ■

research" in the first bill as a permissible reason to possess anthrax. "The bill ended up being as good as we could expect given the current climate," she says.

The ASM has not established how the proposed second law would affect research on anthrax or other agents in the United States, but it should know soon as it is currently notifying its members of the law's provisions. However, Shoemaker says that "the language in the bill is worrisome".

The second bill also requires any laboratory or individual in possession of a select agent to register with the government within 30 days. The ASM estimates that more than 250 labs in the United States possess such microbes or toxins, but Shoemaker claims that many would have difficulty in compiling a comprehensive list in such a short time. "Some labs might not even know what agents they have in their inventory," she says. The ASM plans to push for more time to comply with the new laws at these Senate hearings.

Elizabeth Marincola, executive director of the American Society for Cell Biology, says that her society is still studying the legislation, but agrees that broad restrictions could be harmful. "These things have to be done surgically, not clumsily," she says.

None of the legislation applies to biological specimens that are not listed as select agents, but because the bills refer frequently to "biological agents", the ASM is concerned that their passage may lead to confusion among researchers about what is actually prohibited.

Despite the recent anthrax attacks, routine shipments of biological specimens and reagents have so far been virtually unhindered in the United States. Norman Schwartz, vice-president of the life-science group at Bio-Rad, a California-based manufacturer of materials for biological research, says that his company's orders are going out as usual.

Efforts to manage the transport of dangerous biological agents are even more fraught with difficulty outside the United States. Around the world, germ banks that house agents such as anthrax are already under pressure to increase security, according to bioweapons experts.

But despite tighter security and restrictions on transport, terrorists bent on obtaining anthrax can still turn to the natural world. Anthrax spores survive in soil for decades, and anyone with moderate training in microbiology can isolate them, Collier points out. "You are never going to be able to eradicate them from nature," he says. ■

Legal provision for electronic eavesdropping ignites debate

David Adam, London

Essential surveillance or intrusive snooping?
As governments around the world react to the attacks of 11 September and their aftermath with new anti-terrorism legislation, opinion is divided over whether they are striking the right balance between individual privacy and national security.

The United States, Britain, France and Canada have each prepared such laws, claiming they are needed to track down people who have already committed terrorist acts and to tighten up security. But privacy and other advocacy groups complain that the laws give police and intelligence agencies excessive power to monitor communications, such as use of the Internet, by the population at large.

The USA-PATRIOT Act, signed into law by President Bush on 26 October, for example, does not widen the range of techniques that intelligence agencies can use to intercept and monitor electronic communications — but it makes it easier for the agencies to get legal permission to use the existing techniques.

Under the act, agencies such as the Federal Bureau of Investigation need only show that reading e-mail or sifting through records of electronic traffic held by Internet service providers is "relevant to an ongoing criminal investigation" to obtain this permission from a judge. Previously, they had to convince the judge that the communication to be tapped was linked to actual criminal activity.

In a statement on 11 October, after an initial version of the bill had been passed, Bush said that the new law "respects our constitution" while giving law-enforcement agencies "essential additional tools to combat terrorism and safeguard America against future terrorist attacks".

But some argue that the act lowers the legal barrier to electronic surveillance too far. "There is an obvious risk to privacy every time the standards for government access to personal data are weakened, and in this legislation standards are being substantially weakened, particularly in intelligence investigation," says Jim Dempsey, deputy director of the Center for Democracy and Technology, a Washington-based group that advocates privacy on the Internet.

On 15 October, British home secretary David Blunkett outlined a series of anti-terrorism measures which the government intends to make law by the end of the year. They also include rules to allow easier access to records of electronic communication.

Blunkett's proposals would enable Internet service providers to keep extensive



Listening in: these 'golf balls' in Yorkshire are thought to gather data for British intelligence.

logs of e-mail traffic and Internet use, as they often do in the United States; this is currently prohibited by European data-protection laws. British police already have powers to access such records where they exist.

Similar laws just passed in France mean that records from Internet service providers will be made available to police. And the German Bundestag will debate new security legislation on 7 November, although it is not clear whether it will change the rules safeguarding Internet privacy.

A spokesperson for the British Home Office says that some concerns over the proposed powers are misplaced. "This does not mean that we are going to, or indeed want to, look at people's e-mails," he says. "We're not talking about the content, we're talking about the actual dialogue itself, the record of who communicated with whom."

Internet service providers also hold lists of the web pages (URLs) visited by individuals. Such data will now be more readily available to US and British law-enforcement agencies. British security agencies will only be allowed access to the 'stem' of the URL (such as www.nature.com), but US law enforcers will be able to identify the specific page.

Privacy advocates claim that the changes are equivalent to monitoring which library books someone borrows, and constitute an infringement of civil liberties. But others point out that this type of record can already be sold to commercial organizations to help them target their marketing.

"If companies are storing some of this data anyway for commercial use, then to argue that it's an imposition to provide it under proper authority to law-enforcement agencies strikes me as specious," says James Lewis, director of technology policy at the Center for Strategic and International Studies, a Washington-based foreign-affairs think-tank. ■

Science marshalled to cut bioterror confusion

Rex Dalton, San Diego

Scientists began to play a more prominent role in telling the American people about bioterrorism this week, as the Bush administration responded to heavy criticism of its initial public statements on anthrax.

Top researchers, including John Marburger, the president's newly confirmed science adviser, and Tony Fauci, director of the National Institute of Allergy and Infectious Diseases, joined senior members of the administration in public appearances. As well as assuming a more visible public role in discussing bioterrorism attacks, the researchers took part in top-level meetings in Washington to plan a response to anthrax or other agents that might be directed at the United States.

Until about a week ago, administration officials, such as Tom Ridge, head of homeland security, and Tommy Thompson, the health secretary, often spoke alone, providing conflicting information — and occasional misstatements were blamed on their lack of scientific expertise.

Responding to the change in emphasis, Bruce Alberts, president of the National Academy of Sciences, said that it was "better late than never". Alberts, who was strongly critical of some initial government responses to the bioterror attacks, added that the administration had been "ambiguous about who is in charge" of its scientific assessment of the crisis.

But with Marburger confirmed by the Senate on 23 October as Bush's science adviser and director of the White House Office of Sci-

ence and Technology Policy (OSTP), Alberts and others are hoping for a change.

"What makes sense now is a close partnership between Ridge's homeland security office and the OSTP," Alberts said. And the partnership did not take long to materialize: on 29 October, Marburger appeared alongside Ridge at his daily public news briefing, which was also attended by scientists from the Centers for Disease Control and Prevention (CDC) and the US Army.

But Ridge's office, which was created after the 11 September attacks to guard the home front in the war against terrorism, has only around a dozen senior staff and limited authority over agencies such as the CDC, the Central Intelligence Agency, the Federal Bureau of Investigation and the defence department.

Already, there is evidence of turf battles among these organizations, as they respond to the anthrax attacks. One scientific source says there have been significant breakdowns in communication among some of the agencies.

As the bioterrorism issue develops, concern remains over still-unfilled scientific posts in the administration, including those of director of the National Institutes of Health (NIH) and commissioner of the Food and Drug Administration.

Fauci's emergence over the past week as a prominent spokesman for the administration on public-health matters has fuelled speculation that he may be about to take the NIH position. ■



John Marburger, the president's new science adviser, aims to assume a high-profile role.



Homeland security chief Tom Ridge has started to strengthen links with scientists.

BROOKHAVEN NATIONAL LAB.

R. EDMONDS/AP

Germany warned over threat posed by nuclear reactor

Quirin Schiermeier, Munich

A nuclear watchdog is warning that a German research reactor will continue to

pose a security risk, despite a last-minute plan to cut the amount of weapons-grade uranium to be used at the site.

The FRM-II reactor, housed at the Technical University of Munich in Bavaria, is designed to run on highly enriched uranium (HEU), weapons-grade material that contains a high percentage of the uranium-235 isotope.

Under the compromise, the reactor would be converted to run on medium-enriched uranium (MEU) by 2010. The Bavarian government, which is responsible for running the reactor, defines MEU as containing less than 50% uranium-235.

But the Washington-based Nuclear Control Institute, a non-profit body that monitors worldwide nuclear activity, says that the compromise will do little to reduce the proliferation risk from the reactor.

"MEU is an artificial term," says Paul Leventhal, director of the NCI. "Uranium containing more than 20% of uranium-235

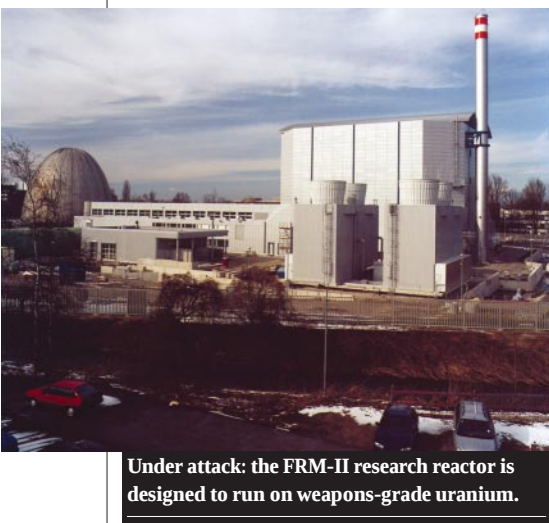
is weapons-grade, and therefore classified as HEU. This agreement means that weapons-grade material will have to be stored at a place without military protection against terrorist attacks."

The NCI says that low-enriched uranium (LEU), which contains less than 20% uranium-235, is the only safe option. But the Bavarian authorities have previously rejected this idea because costly structural changes would be needed for the reactor to use LEU.

Construction of FRM-II was finished in July, but the federal government refused to let it go into operation until the plan to phase out HEU was agreed.

The reactor could enter its eight-month test phase this coming January if, as expected, the government grants the reactor an operating licence next month. The compromise has the backing of the German cabinet, as well as scientists who have booked research time at the facility. ■

♦ www.nci.org



Under attack: the FRM-II research reactor is designed to run on weapons-grade uranium.

Chance encounter hints at planetary origins

Alison Abbott, Munich

Results from the first experiments to be carried out aboard the International Space Station are being analysed — and have



Greg Morfill celebrates the mistake that offered him fresh insight into how planets form.

already sparked debate among plasma physicists about the processes behind the formation of planets.

The experiments, run by a German–Russian collaboration, involved the study of plasma crystals. Plasma is the most disordered form of matter. But order — in the form of crystallization — can be produced by injecting microspheres made of melamine formaldehyde into the system. These become negatively charged amid the electrons and ions that form the plasma, and their electrostatic interactions then cause them to arrange themselves in an ordered way.

Because this effect of the microspheres is disturbed by gravity, scientists designed a series of gravity-free experiments on the space station (see *Nature* **405**, 7; 2000).

At a workshop held in Garching, near Munich, last month, the researchers reported a result obtained purely by chance. On one occasion, a cosmonaut forgot to make the plasma before injecting the microspheres. Without a plasma, the microspheres should remain uncharged and just float about randomly, they say. But instead, the microspheres coagulated almost immediately.

“One hundred thousand particles coagulated in a matter of seconds,” says Greg

Morfill, a director of the Max Planck Institute for Extraterrestrial Physics in Garching, and joint chief investigator of the experiments. “We couldn’t believe what we were seeing at first.”

Uncharged particles would take a day to aggregate through a chance encounter, he says. Further experiments showed that the injected microspheres did in fact carry opposite, attractive, charges, and this produced a millionfold acceleration in the coagulation.

Morfill’s team proposes that, in the absence of plasma, a negative charge is generated on some of the microspheres through the transfer of electrons from the metal of the grid through which the microspheres are injected. These negatively charged microspheres could then cause a redistribution of charge on newly injected, neutral spheres. For a split second, the side of the neutral microsphere closest to the approaching, negatively charged sphere would become positive. Then the particles would collide and shatter, leaving the resulting fragments differentially charged.

Morfill, a theoretical astrophysicist, suggests that this mechanism might help to explain an enduring mystery about how planets form. Scientists believe that planets emerge from the disks of gas and dust that surround new stars, as the swirling dust particles coagulate into ever-bigger particles. But as they aggregate they tend to be accelerated towards the star. The problem for theoreticians is how the particles ever reach the critical size of a metre or so that can resist the gravitational pull of the star. Calculations indicate that aggregation would occur too slowly to overcome the star’s drawing power.

Morfill theorizes that the particles could become differentially charged by the same charge-redistribution mechanism and shattering that were seen in the failed plasma-crystal experiment. This could allow sufficiently fast coagulation to the critical size, he says.

The idea is likely to prove controversial among planetary scientists. Jürgen Blum, of the University of Jena in Germany, is principal investigator of another experiment planned for the space station, ICAPS (Interactions in Cosmic and Atmospheric Particle Systems), which will study the aggregation of dust particles in space. “I’m not sure that there could be an abundance of charge on particles this size,” he says. “But any new experiments that indicate how planet precursors can be formed are very helpful.”

Morfill’s team will have a chance to explore his theory further in another batch of experiments scheduled to take place on the space station early next year.

ALISON ABBOTT

UK changes policy on life insurance

David Adam, London

The British government has suspended the use of genetic test results to determine premiums for life insurance.

The five-year moratorium is an attempt to halt the trend of insurers using data from genetic tests to set policy rates. Such tests can indicate whether an individual is likely to develop certain life-threatening diseases.

Last year, the government gave insurers the right to use a test for Huntington’s disease to set premiums for life and critical illness. This was the first such approval anywhere in the world, but it was based on the fact that Huntington’s is a special case: anyone carrying the gene will develop the degenerative brain disorder.

But in February of this year it was revealed that some insurance companies were using results from three genetic tests that had not been approved — for early-onset Alzheimer’s disease, and hereditary breast and ovarian cancers. The government’s Genetics and Insurance Committee is currently considering whether insurers should be able to use these tests in policy assessments.

Health minister Philip Hunt described the voluntary moratorium as “breathing space” to work with insurance companies to

get things right, but warned that the government would legislate if the industry did not comply.

Insurance companies can still use the test for Huntington’s disease, but only for very large policies, such as life insurance of over £500,000 (US\$730,000). Previously, the test was used for all levels of life insurance. Companies will also be able to use other tests when providing very high levels of cover, but only if the tests manage to get approval from the Genetics and Insurance Committee.



Spotlight on scrapie in hunt for sheep BSE

Declan Butler

Britain is to step up its efforts to discover whether bovine spongiform encephalopathy (BSE) is present in the national sheep flock. Some experts fear that the disease may exist in sheep but that it is mistaken for scrapie, which is endemic in the flock.

David King, the government's chief scientific adviser, announced that the government will test thousands of scrapie cases for the agent that causes BSE. The move, which critics say amounts to a reversal in government policy, follows last month's controversy over whether researchers in another experiment had spent five years accidentally testing cow brains instead of sheep brains for BSE (see *Nature* **413**, 760; 2001).

The new project revives a proposal made three years ago to use a cheap technique that distinguishes between strains of prion protein by revealing their conformations and glycosylation patterns. The test takes just 48 hours.

Until now, UK government experiments on BSE in sheep have relied exclusively on conventional strain typing, which discriminates prion strains well but takes two years to complete because it requires analysis of incubation times and lesion profiles in mice inoculated with sheep extracts. As a result, only 180 sheep brains have so far been tested.

King said he was "confident" that the tests will gather a significant amount of data in a



Fast track: rapid testing of brain samples will help to clarify whether Britain's sheep have BSE.

short time. But he added that, because the test was still being validated by the government's Veterinary Laboratory Agency, it would take "months, rather than weeks to be statistically certain that there is no BSE in sheep".

The test to be used was developed in 1996 by John Collinge of Imperial College School of Medicine in London, a member of the UK government's Spongiform Encephalopathy Advisory Committee (SEAC).

But some scientists are sceptical of the test's ability to distinguish scrapie and BSE. For example, a group at the Institute of

Animal Health near Newbury, Berkshire, has reported that one scrapie strain, CH1641, gives a similar glycosylation pattern to BSE.

SEAC nonetheless said in 1999 that glycoform profiling should be a high priority. Collinge complains that the Department for Environment, Food and Rural Affairs (DEFRA) has not used the test enough to improve its distinguishing power. But Oliver Cattermole, a DEFRA spokesman, says that the validation process is time-consuming, and adds that the test is "not the be all and end all" of sheep BSE research. ■

Middle Eastern promise as synchrotron wins approval

Quirin Schiermeier, Munich

The scientific commission of the United Nations Educational, Scientific and Cultural Organization (UNESCO) has formally endorsed the creation of a synchrotron radiation facility in the Middle East.

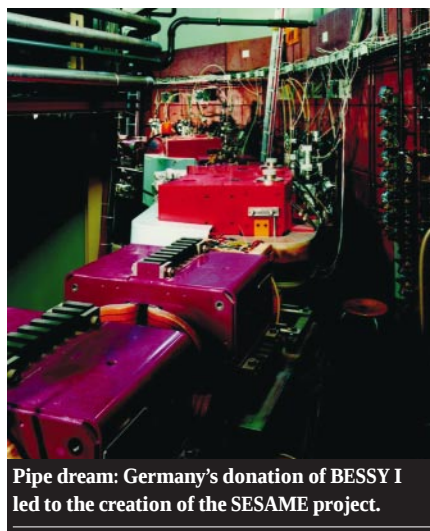
The facility, planned for construction near Amman in Jordan, will operate under the auspices of UNESCO, whose general assembly, currently sitting in Paris, is expected to approve the plan on 3 November.

The International Centre for Synchrotron-light for Experimental Science and Applications for the Middle East, known as SESAME, is a joint project by Armenia, Cyprus, Egypt, Greece, Iran, Israel, Jordan, Morocco, Oman, the Palestinian Authority and Turkey. The idea was born in 1999, when Germany offered to donate its dismantled BESSY I synchrotron to the Middle East.

The agreement with UNESCO is a big step forward for the project. UNESCO will provide the institutional framework for hiring staff and for all contractual arrangements between member states, as it did 50 years ago for CERN, the European

particle-physics laboratory in Geneva.

Jordan will pay US\$6million for a new building, and the European Union and the International Atomic Energy Agency have promised to support the installation and modernization of the German synchrotron.



Pipe dream: Germany's donation of BESSY I led to the creation of the SESAME project.

The SESAME member states must cover the rest, including the running costs.

Synchrotrons produce high-quality X-rays by circulating electrons around a ring at high speed. Basic researchers from many disciplines use them to study the molecular structures of organic and inorganic matter.

"In the current political situation, scientific exchange in the Middle East is even more important than before," says Howard Moore, a spokesman for UNESCO's science division. "This will become a state-of-the-art machine," says Dieter Einfeld, technical director of SESAME, who is overseeing the training of 20 scientists from the region to build and operate SESAME.

Operation could begin within as little as three years, says Herwig Schopper, former director general of CERN and head of SESAME's interim council. But the synchrotron must first be shipped to Jordan and upgraded to achieve higher energies and the short wavelengths required for structural biology. The storage ring is to be lengthened to 120 metres in circumference, and equipped with magnets to enhance its performance. ■

Intel founder chips in with 'largest ever' gift to academic world

San Francisco The California Institute of Technology (Caltech) has pocketed US\$600 million from microchip billionaire Gordon Moore and his wife Betty, an award that the institute claims is the largest donation ever made to an academic centre.

Half of the donation came directly from the couple themselves, and the other half was channelled through the Gordon and Betty Moore Foundation, a San Francisco-based grant-making organization established by them a year ago.

Gordon Moore earned a PhD in chemistry from Caltech in 1954 and was one of the founders of computer-chip manufacturer Intel in 1968. He has made several previous donations to Caltech and is a long-standing member of its board of trustees.

♦ www.moore.org

All goes swimmingly for puffer-fish genome

London A draft sequence of the puffer-fish genome has been completed by an international consortium in just one year.

Fugu rubripes has a repertoire of genes that is very similar to that of humans, but it crams them into a genome that is one-eighth the size of ours. The similarities should help to speed up the discovery of genes and their key controlling sequences in the human genome. "These sequences are some of the most exciting targets for therapeutics," says Greg Elgar of the UK Human Genome Mapping Project Resource Centre near Cambridge, a member of the consortium.

The draft sequence could also help to settle disputes over the total number of genes in the human genome, as the compact

nature of the *Fugu* genome makes counting genes relatively easy. The rough sequence, which covers 99% of *Fugu*'s DNA, contains between 35,000 and 40,000 genes. The International Fugu Genome Consortium plans to publish an analysis early next year, and will make all sequence information freely available.

♦ fugu.jgi-psf.org

NASA in orbit over Mars Odyssey success

Washington NASA officials breathed a sigh of relief last week as the 2001 Mars Odyssey spacecraft successfully entered into martian orbit. The agency lost contact with Odyssey's two predecessors on its Mars exploration programme — a lander and an orbiter, which were valued at a total of US\$290 million — after they arrived at the red planet.

After entering its orbit around Mars, Odyssey sent back radio signals reporting its success, and indicated that its 'aerobraking' technology was working. The technology, developed for interplanetary travel and tested on NASA's Magellan mission to Venus in the early 1990s, will allow Odyssey to use drag from the martian atmosphere over the next three months to slowly bring itself into a lower orbit 400 km above the planet's surface.

Spectrometers and imagers on the \$300 million craft will then record the geology and climate of the planet, and look for water. The spacecraft will also provide communications relays for future US and European lander missions scheduled for launch in 2003.

Europe sets target date for Kyoto Protocol

Munich The European Commission has asked the 15 member states of the European Union (EU) for a ratification date of 14 June 2002 for the Kyoto Protocol on reducing greenhouse gases, and has put forward proposals on how this should be done. Details of the protocol, including the right

to trade emissions targets with less industrialized countries, will be discussed this week as climate-change talks resume in Marrakech, Morocco.

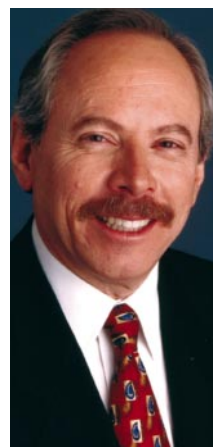
Once adopted, the protocol will commit the EU to cut its carbon dioxide emissions by 8% from the 1990 level by 2012. Under the EU proposals, reductions for member states vary sharply according to how easy it is for the states to cut their emissions. Germany and Britain are asked to reduce their emissions by 21% and 12.5%, respectively; France need not make any change from its 1990 levels; and Greece is allowed an increase of 25%.

"We hope this proposal will encourage

other countries to ratify soon so that the protocol can enter into force before the World Summit on Sustainable Development in Johannesburg in September next year," says Margot Wallström, the European commissioner for the environment. The proposal has been forwarded to all member states for approval.

Neuroscientist lined up for top job at AAAS

Washington Alan Leshner is to take over the reins of the American Association for the



Alan Leshner: stepping in as chief executive officer at the AAAS.

Advancement of Science (AAAS) when the current chief executive officer Richard Nicholson retires next month after 12 years in the job.

Neuroscientist Leshner, who is 57, has been head of the National Institute on Drug Abuse at the National Institutes of Health since 1994.

The AAAS is best known as the publisher of the weekly journal *Science*, and for its

annual meeting dedicated to the public understanding of science. The non-profit-making society, which has an annual budget of approximately US\$80 million, also supports programmes aimed at improving science education and influencing science policy.

♦ www.aaas.org

Imaging network gives brain scientists a head start

San Diego The University of California, San Diego, has won a \$20 million grant from the US National Institutes of Health to build a high-performance computer network for neuroscientists.

The network will allow researchers at different universities to share digital images of the brain generated using different techniques, such as magnetic resonance imaging and *in vivo* three-dimensional microscopy. Such techniques allow the shape of functioning brains to be visualized.

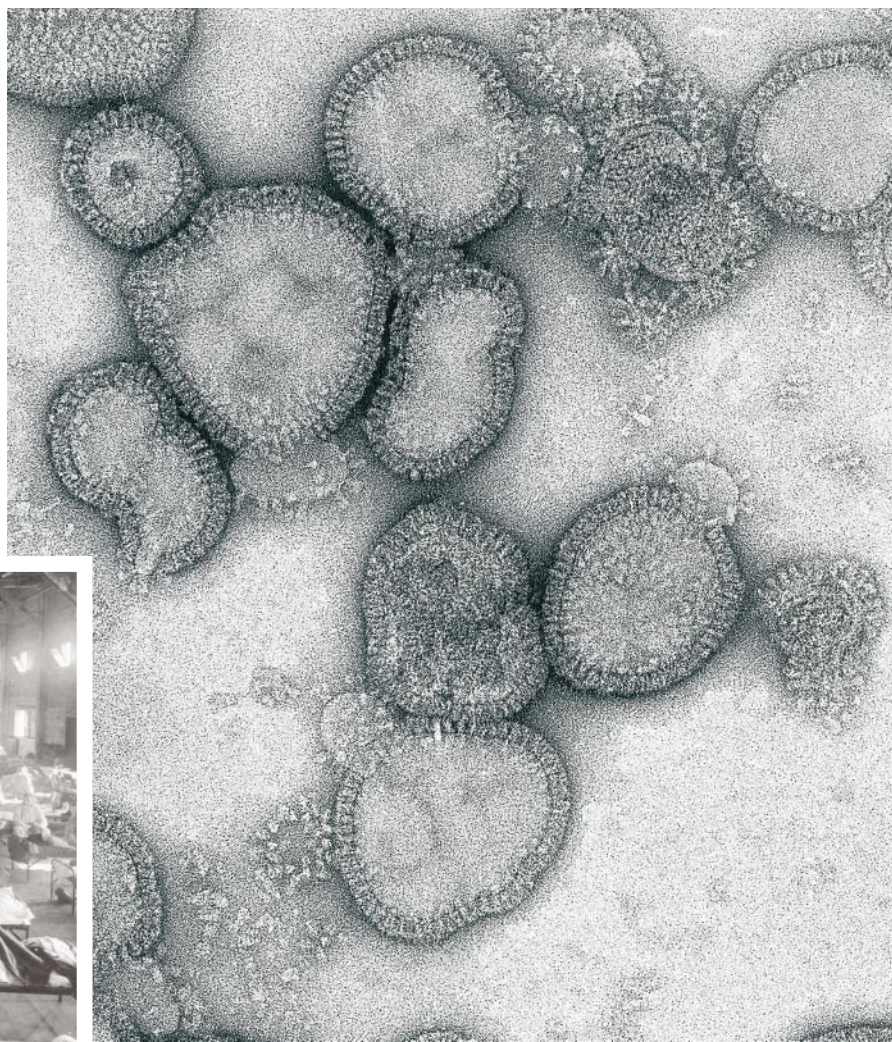
The project, known as the Biomedical Informatics Research Network (BIRN), will initially be made up of two sub-projects. Mouse BIRN will allow sharing of data on mouse models of multiple sclerosis and dopamine disorders. The other sub-project, Brain Morphology BIRN, will contain data from human studies, with an initial focus on depression and Alzheimer's disease.



On a plate: the puffer-fish sequence could aid attempts to understand the human genome.

The flu HQ

Leading virologists plan to thwart emerging strains of influenza by creating a global laboratory to keep tabs on this ever-changing virus. Alison Abbott reports.



L. CALDER/NIMR

AP

Can streamlining the analysis of flu viruses (above) avoid a repeat of the 1918 pandemic?

During the Northern Hemisphere's coming winter, tens of thousands of people may die of influenza. It is a sobering figure, but a mere fraction of the half-million killed in 1968, the last time a flu pandemic swept the globe. And compared with the 40 million who perished in the notorious pandemic of 1918, it is a drop in the ocean.

But sooner or later, another strain of flu virus will emerge with the potential to wreak comparable havoc. When it does, minimizing loss of life will depend on rapidly identifying the threat and producing appropriate vaccines. And that, says an influential group of leading flu researchers, means streamlining the existing monitoring system and creating a new global influenza laboratory. "We need to find a better way to predict the inevitable pandemic, and control it," argues virologist Robert Webster of the St Jude Children's Research Hospital in Memphis, Tennessee.

The proposed global laboratory would build on, and partially replace, the current worldwide flu-monitoring system. This was set up in 1948 by the World Health Organization (WHO) to track circulating strains of the

virus in both hemispheres. It comprises more than 100 national centres in 83 countries plus four large 'collaborating centres' in London, Melbourne, Tokyo and Atlanta, Georgia.

Flu is an RNA virus with eight genes. Vaccines to combat it must be updated regularly because the virus frequently changes its appearance in an effort to elude its hosts' immune systems. It does this by mixing and matching varieties of two of its genes. One encodes the protein haemagglutinin, which helps the virus enter host cells; the other generates the enzyme neuraminidase, which allows the virus to burst out of host cells once it has replicated.

Taking the strain

At present, the national centres collect and screen samples from about 175,000 of the billion or so cases of flu that occur worldwide each year. They classify the viruses into two major subtypes, A and B. Because A is a more variable subtype than B, they also document its haemagglutinin and neuraminidase proteins. Most of the flu viruses circulating this year, for instance, are A(H1N1), A(H3N2) or B.

The national centres send a representative fraction of the more common strains to one of the collaborating centres for detailed analysis, often including sequencing of the haemagglutinin and neuraminidase genes. All unusual strains are also sent for further analysis. In addition, the national centres deposit epidemiological information into a WHO database set up in 1997 and recently made available online as FluNet, which collates data on disease incidence and circulating strains.

In February and September, scientists from the four collaborating centres meet to decide the composition of the next vaccine. Manufacturers have about six months to prepare the 100 million or more shots that are delivered in each hemisphere's flu season.

Most experts agree that the system works well — but not well enough to be sure that it can react to an emerging pandemic with the necessary speed. Taking advantage of high-throughput gene-sequencing technologies, leading flu specialists — including representatives of the four WHO collaborating centres — want to establish a central global facility to type and sequence samples submitted by national centres. "This is the lab of the future:

AP highly automated, very centralized and very efficient," says Alan Hay, director of the WHO Collaborating Centre for Reference and Research on Influenza at the National Institute for Medical Research in London.

Rapid response unit

At present, the four collaborating centres between them produce sequence information on only 4,000 or so samples each year. Under the global laboratory proposal, outlined in September in *Science* (293, 1729; 2001), this would be increased by at least an order of magnitude. Results would be made available to those submitting samples within days — currently, this can take several weeks. The national centres would increase their gathering of epidemiological data, which would be submitted to the central lab.

With the swifter flow of greater volumes of data, up to a month could be gained between the identification of new circulating viruses and the start of vaccine development, say the plan's proponents. But Klaus Stöhr, an official responsible for flu at the WHO's Geneva headquarters, notes that the biggest delay — of between six and eight months — is in vaccine manufacture. "A month or two will not make up this lag," he says.

The global lab should bring other gains, however. Good information about mutations in the neuraminidase gene, for instance, would help keep track of the emergence of resistance to drugs that work by inhibiting the enzyme, such as Relenza (zanamivir).

Perhaps most significantly, the global lab should accelerate understanding of the virus's biology — which would help explain why some strains of flu are more dangerous than others. This knowledge could allow virologists to predict whether newly emerging strains have the potential to cause a pandemic.

Even in the absence of a pandemic strain, scientists need to get a better handle on why flu is more virulent one year than the next, says Michèle Aymard, head of the National Reference Centre for Influenza in Lyon: "Flu is so common and we know so little about it." In the winter of 1992–93, for instance, Aymard observed unusually high rates of hospitalization in France, as well as outbreaks in elderly

people who had been vaccinated, without being able to provide a scientific explanation.

Nancy Cox, head of the WHO collaborating centre at the US Centers for Disease Control and Prevention in Atlanta, is convinced that widespread sequencing of genes other than those for haemagglutinin and neuraminidase will reveal important virulence factors. "We need to find other molecular signals for changes in virulence for which we do not have good markers," she says.

One example could be the *NS1* gene, which Peter Palese, a flu virologist at the Mount Sinai School of Medicine in New York, suggests could promote virulence because the protein it encodes neutralizes the effects of interferon, an antiviral protein made by the immune system. He suspects that mutations in this gene may explain why the strain of virus that caused the 1918 pandemic was so extraordinarily virulent. "A protein that targets the host's defence response could provide a logical explanation," says Palese. The chance to compare *NS1* mutations with epidemiological data would put the theory that the gene is a key virulence factor to the test.

Haemagglutinin itself is also known to be a powerful virulence factor. The major twentieth-century pandemics — caused by an H1N1 virus in 1918, H2N2 in 1957 and H3N2 in 1968 — were each associated with a subtype of the haemagglutinin gene to which people had not previously been exposed.

Exchange rates

It is widely believed that these pandemic viruses arose because of gene reassortment in which a virus from an animal — birds, pigs and horses all harbour flu — co-infected a host cell with a human virus and an exchange of genes occurred. "Observing the frequency of reassortment of genes in strains analysed by the global laboratory will be important to see the processes involved in generating pandemics," says Cox.

The global laboratory would also monitor strains of flu in animals. The importance of this was underlined in late 1997, when an H5N1 strain of flu crossed over from poultry to people in Hong Kong and caused several deaths. In the event, it was not the start of a pandemic — but the threat of a devastating virus suddenly emerging from birds or other animals remains. "The identification of novel influenza subtypes in animals may turn out to be directly predictive of a human pandemic and thus allow vaccines to be manufactured in advance of a major outbreak," says Jeffrey Taubenberger, a virologist at the US Armed Forces Institute of Pathology in Washington.

Although the global laboratory plan has strong support among independent flu virologists and at the existing collaborating centres, some scientists at the national centres remain unconvinced. "There are better and cheaper ways to improve the flu surveillance system," argues Reinhard Kurth, head



Nancy Cox: wants more genetic information.



Pandemics could be thwarted by spotting new flu strains in animals, says Jeffrey Taubenberger.

of the Robert Koch Institute in Berlin, which is home to one of Germany's national centres. He points to the WHO's current efforts to improve communication between national centres, through initiatives such as extending FluNet and establishing a new helpdesk in Geneva to answer queries from the centres.

Kurth also worries that scientists at the national centres will lose their enthusiasm if their jobs are reduced to posting samples for analysis elsewhere. But Hay argues that this will not happen: "Really it will liberate them from the routine — the truly scientific work will still need to be done."

Moving the plan forward will require money — around US\$50 million for the first five years, to build, equip and run the global laboratory. Although the plan envisages that WHO would continue its coordinating role, the agency does not have the necessary money to fund the lab. So the plan's proponents are looking to philanthropic organizations such as the Bill & Melinda Gates Foundation.

Scott Layne, a physicist-turned-physician at the University of California, Los Angeles, who is one of the driving forces behind the initiative, is confident that it "could save untold numbers of lives". If so, the global laboratory concept might be applied to other diseases. And, in the current climate, it may also be examined in the context of monitoring bioterrorist incidents.

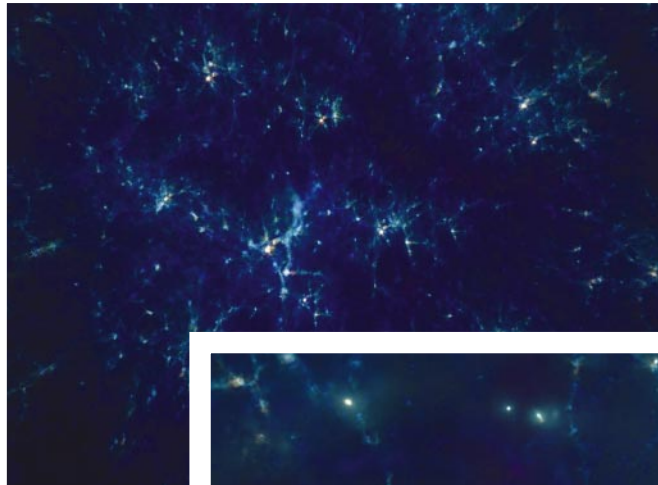
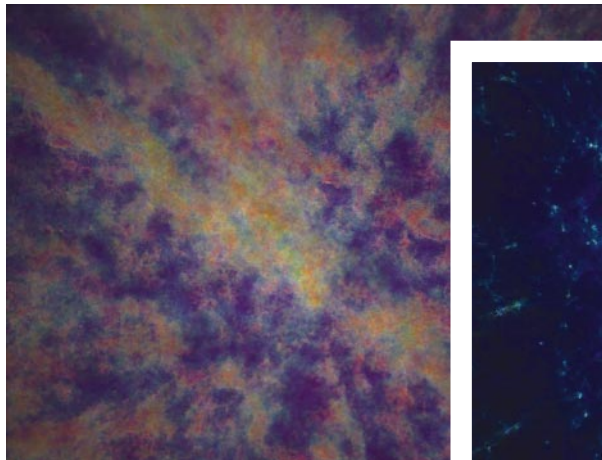
Alison Abbott is *Nature's* senior European correspondent.

WHO flu factsheet • www.who.int/inf-fs/en/fact211.html

FluNet • oms.b3e.jussieu.fr/fluNet

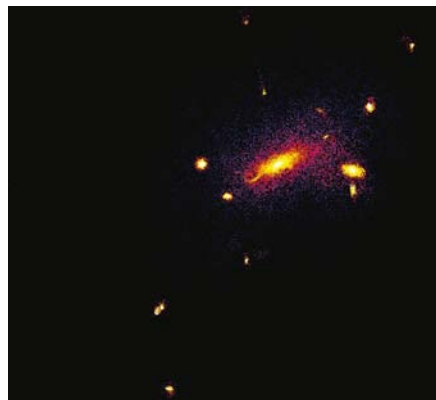
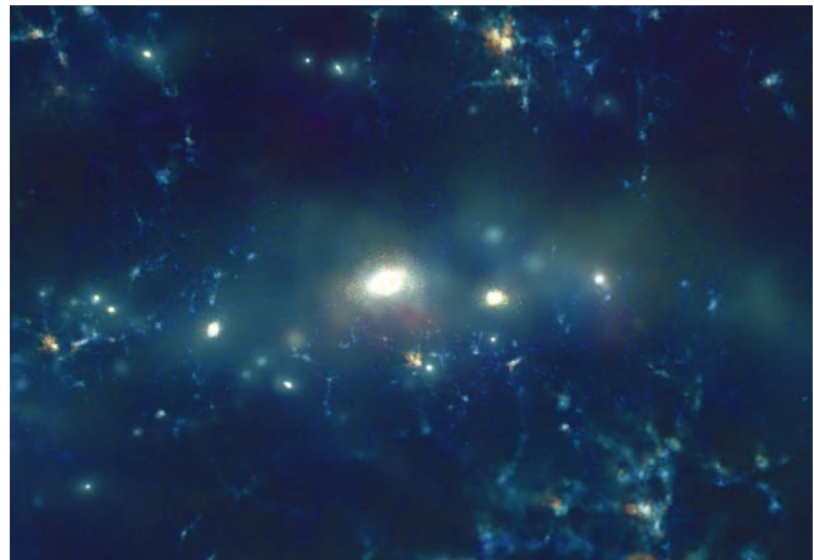


Flu from animals, such as poultry, can infect people, as happened in Hong Kong in 1997.



Exploring a virtual universe

Cosmologists have already created entire universes within computers. Now astrophysicists are focusing on the fine details of asteroid collisions and supernovae. Govert Schilling investigates.



Heckling is not something scientists expect when they present results to their peers. But when Simon White revealed the image above — which shows a cluster of galaxies — during a meeting on galaxy formation held in Leiden last May, a member of the audience shouted out: “Is that an observation or a simulation?”

Such a question would have been unlikely 20 years ago. Then, many researchers viewed computer simulations with suspicion. But in the 1980s, computer models of how large-scale structures such as galaxy clusters evolved began to produce impressive results. More recently, new computing techniques have allowed smaller-scale processes, from

asteroid collisions to supernova explosions, to be modelled with unprecedented accuracy. White’s image was a simulation, but the fact that the question needed to be asked shows just how successful computational studies have become.

The early simulations helped to shed light on some of the mysteries of the Universe. Galaxies and galaxy clusters evolved from the soup of particles and radiation present after the Big Bang. The laws that control this evolution, such as Newton’s laws of motion, were well understood. But other factors, such as the role of particles of cold dark matter (CDM), were less well defined. Researchers suspected that CDM made up a significant proportion of the mass of the Universe, and so affected the way it had evolved. But because the particles do not emit light or electromagnetic radiation, they had never been observed directly.

To test different ideas about CDM’s role, cosmologists needed an experimental universe. Tweaking the initial conditions of virtual universes, or the laws that govern them, changes the way in which they evolve. By searching for the simulations that resulted in universes similar to our own, researchers were able to test different theories.

From the early 1980s onwards, computers were powerful enough to simulate useful vir-

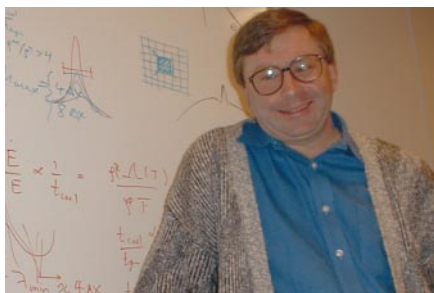
Growth factor: a cosmology simulation showing galaxy clusters forming from cosmic dust.

tual universes. Researchers found that large-scale structures, such as clusters of galaxies, would only evolve if their simulations included large amounts of CDM. The studies gave cosmologists confidence in their assumptions, and most now believe that CDM makes up about 95% of the mass of the Universe.

In the dark

Despite these successes, many issues in cosmology and astrophysics, such as the formation of individual galaxies, remain difficult to simulate. “It’s relatively simple to trace the distribution of dark matter in the Universe,” says White, who directs the Max Planck Institute for Astrophysics in Garching, Germany. “But with luminous matter, there’s a lot more physics involved, such as star formation processes and supernova explosions.”

Fortunately, researchers such as White will soon be able to call on a new generation of supercomputers. This July, the Virgo Consortium of European researchers inaugurated its new Cosmology Machine, based at the University of Durham in England, which can perform 228 billion calculations per second (228 gigaflops). In the United States, a con-



Simulations are no longer a 'black art' for cosmology and astrophysics, says Mike Norman.

sortium headed by the San Diego Supercomputer Center at the University of California, San Diego, and the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, has also recently announced plans to build a system that can do a staggering 13.6 trillion calculations per second (13.6 teraflops). The system will be used by researchers in several fields, but US cosmologists and astrophysicists expect to be allocated some time on it.

Using these new machines, it may be possible to fix some of the problems with simulating galaxy formation. All large galaxies, for example, appear to be embedded in a halo of CDM. The halo's structure can be inferred by measuring the effect its gravitational field has on the motion of the galaxy it surrounds. But the results of observations of this effect do not tally with those produced by simulations.

All for one

Both of the new computing systems are examples of a recent trend in computer design. By linking networks of less powerful machines together to form clusters of processors, researchers are creating systems that can match the speed and power of bigger machines, but at a fraction of the cost. The NASA scientists who created the first such network in 1994 christened their cluster Beowulf — the title of an eighth-century English poem — and the name has stuck.

The San Diego and Durham systems are examples of relatively expensive Beowulf systems. More usually, such clusters are deployed by those with smaller budgets, and some of these less expensive set-ups are now being used to tackle the smaller-scale processes that have yet to be simulated successfully.

"Beowulfs are the platform of choice for astronomy departments that don't have unlimited money," says Derek Richardson, an astrophysicist at the University of Maryland in College Park. Together with colleagues at the University of Washington in Seattle, where he worked until last year, Richardson has used Beowulfs to model the way asteroids behaved in the early Solar System¹.

Small bodies, such as asteroids and comets, grouped together over millions of years to form planets. Richardson simulated

collisions between one type of asteroids and showed that only rarely — in around 30% of cases — do they stick together when they collide. His work suggests that collisions between equal-sized bodies played only a minor role in planetary formation, and that the gravitational pull of large asteroids on smaller bodies is the more important process.

But the Beowulf concept does have its drawbacks. Beowulf software has to be designed so that it can run on a large number of processors at once. A grid-like code, where space is divided into tiny sub-regions, is ideal. To model the fluid dynamics of the gases inside the Sun, for example, the Sun is divided into a series of cells. Each individual processor simulates the behaviour of gases in a small number of cells by modelling the physics taking place inside them and assessing the influence of their immediate neighbours.

Such problems lend themselves to Beowulf systems because each spatial cell is influenced only by neighbouring cells, and the amount of communication between processors is limited. But in other simulations, things are not so simple.

Erik Asphaug of the University of California, Santa Cruz, has used a low-cost Beowulf system to model the collision between the young Earth and a second proto-planet which led to the formation of the Moon². He divides the system up into a series of 'test particles', groups of which are simulated by individual processors. But the most important force in the system — gravity — acts across all the test particles, so each processor has to communicate with all of the others. This slows the simulation down. Asphaug admits that he is envious of people who have code that can be run in parallel. Richardson's asteroid simulations, in which gravity is also very important, suffer from similar problems.

Globular clusters — spherical groups of

stars that lie within, or orbit, galaxies — are also difficult to simulate using Beowulf clusters. A Beowulf network could, in principle, be used to model them. But, unlike the rocks in Asphaug's simulations, stars within the clusters cannot be lumped together as single test particles. Stars of different masses, for example, evolve in different ways. And because clusters contain hundreds of thousands or even millions of stars, simulations of them require an equivalent number of test particles.

Fruitful arrangement

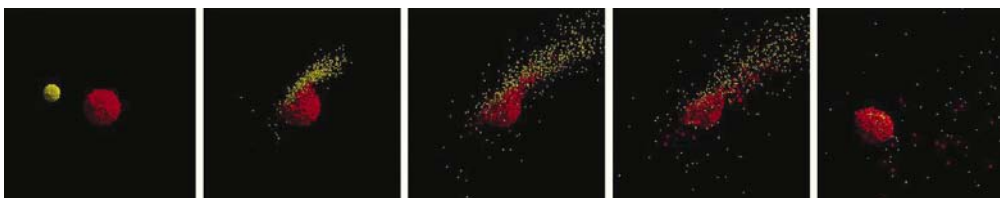
Beowulf networks of millions of processors are not feasible, so researchers modelling globular clusters at the Institute for Advanced Study in Princeton have taken a different approach. Rather than writing software to run on general-purpose machines or networks, they are using machines that are hard-wired to solve the equations describing the gravitational interactions. The machines, known as GRAPE-6s, were designed by researchers at the University of Tokyo and, with a peak performance of tens of teraflops, were last year awarded the title of fastest computers in the world.

The Princeton team says that it will soon be able to run simulations of a few hundred thousand stars. The group plans to study how conditions in the early Universe influenced the evolution of globular clusters. For example, it hopes to explain why some clusters contain many millisecond pulsars — binary stars in which one member of the pair has sucked in matter from its companion star — whereas others contain only a few such systems.

But GRAPE-6 users can forget about using them to play games during their lunch-break. The machines are designed to tackle specific problems, and cannot be used for anything else. For some researchers, this narrow focus is a disadvantage. "For special



Starring role: a GRAPE-6 machine (above), custom-built to model globular clusters (left).



Crash course: Derek Richardson (left) has modelled colliding asteroids (above).



areas GRAPEs are great," says David Clarke, a computational astrophysicist at Saint Mary's University in Halifax, Canada. "But many areas in astrophysics are messy, so you really need a general-purpose machine."

For some of these messy problems, researchers are turning to individual supercomputers that can be adapted to simulate a range of different physical processes. The new 5-teraflops machine at the National Energy Research Scientific Computing Center, run by the Lawrence Berkeley National Laboratory in California, is one such example. Named after the late nuclear physicist Glenn Seaborg, the Nobel laureate and former associate-director at the laboratory, the machine is the most powerful non-classified, general-purpose supercomputer in the world.

Seaborg is currently being used to study supernovae, which are among the most physically complicated events in the Universe. The explosions involve extreme conditions, including very high densities and temperatures, as well as phenomena that are not well understood, such as strong magnetic fields and the production of gravitational waves. They also evolve extremely fast.

"Many years ago, we started with crude one-dimensional simulations," says Adam Burrows of the University of Arizona in Tucson, one of a team of 12 astrophysicists working on the supernova study. "But one-dimensional supernovae don't seem to explode." Improved two-dimensional simulations have shed light on some processes, such as convection currents in the exploding star. Burrows' collaboration, as well as a rival team at the Oak Ridge National Laboratory in Tennessee, now

plans to use three-dimensional versions to explore the problem in depth.

"We'll take existing stellar evolution models and try to explode them," says Burrows. He plans to run models of stellar evolution until the simulations reach the end of the stars' life. At this point, a real star would start to collapse in on itself. By adding extra detail to the models, such as better descriptions of the nuclear processes within the star, he hopes to create simulations that do the same. The team will then analyse details of the virtual explosions and compare them with data from real supernovae.

In the process, the researchers hope to shed light on some of the unknowns in supernova research, such as the processes that control the production of neutrinos—subatomic particles with very low mass which are generated in the explosions. Neutrinos from explosions close to our Galaxy, such as from the supernova 1987A, can be detected on Earth and linked to their source. Researchers expect similar explosions to occur in our Galaxy every few decades or so. Different types of supernovae probably produce neutrinos in different numbers and with different energies, so better knowledge of these factors could help to classify future explosions.

Planet builders

Tom Quinn of the University of Washington in Seattle is using a Cray supercomputer to study another little-understood issue—that of how the Earth got its supplies of carbon and water. Like other planets, the Earth formed when solid matter in the disk of gas and rubble left over from the Sun's formation grouped together over millions of years. But the part of the disk in which the Earth formed was too close to the Sun for water or carbon to have existed in solid form.

Some researchers have suggested that asteroids containing carbon and water collided with the Earth after being knocked out of orbits further away from the Sun. But initial results from Quinn's simulations suggest that there was little transfer of matter between different orbits in the middle stages of planet formation, around 4.5 billion years ago³. This lends support to the alternative view that water and carbon arrived when the Earth was bombarded by comets later in its life, around 3.8 billion years ago.

Groups that do not have a Seaborg or a Cray of their own might still have the chance to access one, thanks to emerging techniques for sharing computing power. Clarke and col-

leagues at Saint Mary's have recently bought a 10-gigaflops computer. Government funding was conditional on a fraction of the machine's computing time being made available to other Canadian science institutions. This will be done through a network that links some of the country's most powerful computers. Tasks sent to the network are distributed over whichever machines are free at the time, and users can select the type of machine most suited to their simulation.

Clarke's new component for this network will be used by researchers at the new Institute for Computational Astrophysics, based at Saint Mary's. Funding for the institute has been secured, and Clarke says he aims to have five or six staff in place by this time next year. The precise direction of the new team has yet to be decided. "We can't cover the entire field," says Clarke. "We might adopt a theme, such as the origins of stars and galaxies or the life cycle of stars."

Simulation stimulation

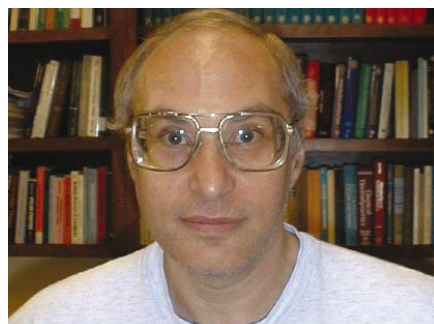
The flurry of fresh activity at Saint Mary's and elsewhere illustrates how important simulations have become. More powerful computers have undoubtedly played a role. But for cosmologists, better observational data—such as the precise measurements of the cosmic microwave background, the radiation that dates from just half a million years after the Big Bang—have also been vital. "These data have really pinned down the cosmological parameters," says Mike Norman of the University of California, San Diego. "A few years ago, you had to run numerous simulations with different parameters. Now you can just take one model and compute the consequences in great detail."

But such success can bring problems. As simulations become ever larger, finding the human resources to interpret them is becoming difficult. "Post-processing and analysing the huge amounts of computer data is very person-intensive," says White. "And it's almost impossible to automate." There is also a need for new software. Computers are getting faster, says Richardson, but computer algorithms have to keep pace.

Such issues were probably the least of the worries for researchers performing the early simulations. "It used to be regarded as a black art, performed by people who possessed specific tricks," says Norman. Many theoreticians refused to believe the results of simulations unless they had been verified by other means.

Now ideas for new observational projects often stem from computer models, showing just how trusted they are. For many in the field, simulations have joined observation and theory as the third pier of astronomy. ■

Govert Schilling is a writer in Utrecht, the Netherlands.



Big bangs: Adam Burrows plans to run 3D simulations of supernova explosions.

1. Leinhardt, Z. M., Richardson, D. C. & Quinn, T. *Icarus* **146**, 133–151 (2000).
2. Canup, R. M. & Asphaug, E. *Nature* **412**, 708–712 (2001).
3. Richardson, D. C., Quinn, T., Stadel, J. & Lake, G. *Icarus* **143**, 45–59 (2000).

Vanishing websites are the weakest link

Sir — Referencing websites for supplementary data and online applications is increasingly popular and acceptable in today's scientific publishing. Yet my recent experience of testing the validity of web links referenced in papers published in 1998 supports the view that many websites are valid for only a short time.

I used PubMed to search for papers using the string "http", and found about 300 citations. Of the 265 web links that appeared in abstracts, only 150 (57%) are still valid; 97 (36%) are no longer available and 18 (7%) redirected me to other URLs. Although these statistics are biased towards biomedical research and might not reflect other disciplines, they are discouraging.

The root of the problem is the formidable challenge of maintaining a website indefinitely. Many websites are stored on local servers, managed by system and network administrators who are not necessarily aware of the need to maintain links cited in publications in unaltered form. Other problems abound: for example, individual researchers might lose

data in computer breakdowns, and the ownership of Internet domain names is often out of their control.

Authors must therefore take great care when citing web links. They must ensure that any website they create to supply additional information will be maintained indefinitely before referencing it in their publications. If authors need to reference websites belonging to other organizations or individuals, they should try to use only those that have been cited in other publications, as these sites are usually better maintained. Authors should redirect outdated web links to current URLs.

Publishers should encourage authors to deposit their supplementary information in databases that are maintained carefully by the publishers. Many publishers now use the Digital Object Identifier (DOI) system to guarantee linkage, but this does not solve the problem of links to supplementary information in various academic or other informal websites. Further, DOI is relatively expensive for academics to set up on their own websites.

Given the convenience of the Internet, more and more information will be stored digitally as time goes on and URL references will become increasingly common. The

lifetime of those references will determine how often we face a missing link.

Joseph Cheung

Department of Genetics and Genomic Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

Nature's policy on supplementary information has always been to host such material on its own website so that its integrity is maintained indefinitely. See www.nature.com/nature/submit/gta/index.html#4.10 — Editor, Nature.

Give and take the lead

Sir — Your News report "Stanford gift scaled back over federal stem-cell policy" (*Nature* **413**, 5; 2001) notes that Jim Clark, the founder of Netscape, has retracted his \$60-million pledge to support a biomedical research centre at Stanford University, in retaliation for President Bush's decision to limit support for stem-cell research. Clark has wasted an opportunity to make good on the boldness inherent in the tradition of strategic private philanthropy.

Private philanthropy, at its best and most influential, is about doing what the government will not do because it is too controversial, risky or unpopular. The Rockefeller Foundation supported population research when timid governments would not. The Parkinson's Disease Foundation nurtured fetal cell transplantation research. The Aaron Diamond Foundation led the way supporting AIDS research.

The federal government must answer to the diverse interests, beliefs and desires of the public. Clark is spending his own money and answers only to himself. Government institutions can have their hands tied by bureaucracy and political expediency. Clark wisely pledged his support to a private institution free from such restraints. The bold response to President Bush's plan shouldn't be withdrawing money — it should be spending it!

So much of the current stem-cell debate is focused on pie-in-the-sky promises or nit-picking technicalities. Let's use private money to fund the research, let's get the data out there, and let's use the scientific findings to guide policy debates. Holding research hostage to a financial game of chicken accomplishes nothing. Clark could have made good on his pledge, upped the ante and challenged others to follow. Taking a stand and mobilizing the vast private wealth of this country — that would be a brave and daring manoeuvre.

Susan M. Fitzpatrick

James S. McDonnell Foundation, 1034 South Brentwood Boulevard, Suite 1850, St Louis, Missouri 63110, USA

Recognizing risks and potential promise of germline engineering

The consequences of mitochondrial DNA manipulation are unknown, but sufferers of some diseases might benefit.

Sir — Your informative News Feature on germline engineering, "Biology's last taboo" (*Nature* **413**, 12–15; 2001), suggests that the presence of donated mitochondria in a human embryo created with donor cytoplasm does not constitute germline engineering. Specifically, the feature states: "Mixing mitochondria does not risk damaging important chromosomal genes, nor could it endow children with eugenic traits." Gregory Stock of the University of California, Los Angeles, is quoted as saying that applying the term 'germline engineering' to this situation is "playing games with words".

Unfortunately, we know that devastating human diseases — heritable through the female line — can be caused by mutations occurring in the handful of genes carried on the mitochondrial genome and required for oxidative phosphorylation. It is not true to suggest that harm to the germline cannot come from mitochondrial manipulation. It's less clear whether mitochondrial DNA manipulation could result in inheritance

of positive traits, but one might speculate that the rate or efficiency of oxidative phosphorylation could affect highly energetic tissues, say of muscle and nervous system, with unknown consequences.

You mention the work of Jacques Cohen and colleagues, using donor cytoplasm and mitochondria to 'rejuvenate' the eggs of women with fertility problems. This is not inconsequential to women of child-bearing age who have been diagnosed with a mitochondrial DNA disease. A process such as this 'rejuvenation' may hold the possibility of allowing these women to bear children who do not carry the mutant mitochondrial DNA in their genomes. Thus it represents exactly the same promise of germline engineering discussed in relation to chromosomal DNA defects. This is not to support or condone the work conducted by Cohen, but merely to point out its significance.

Sharon E. Hesterlee

Muscular Dystrophy Association, 3300 E. Sunrise Drive, Tucson, Arizona 85718, USA

Counting on science

Scientific uncertainties relating to the social issues of today.

**Damned Lies and Statistics:
Untangling Numbers from the
Media, Politicians and Activists**

by Joel Best

University of California Press: 2001. 199 pp.
£13.95, \$19.95

**Fragile Science: The Reality
Behind the Headline**

by Robin Baker

Macmillan: 2001. 272 pp. £15.99

Robert M. May

Where science entwines with public policy and opinion, things are often complicated — or certainly more so than most people seem to realize. This is the common theme tackled by Joel Best and Robin Baker, but using very different approaches.

Best begins by observing that: “Our knowledge about society tends to be ‘softer’ than our knowledge of the physical world. Physicists have far more confidence in their measurements of the atomic weight of mercury than sociologists have in their descriptions of public attitudes toward abortion.” He then describes how social scientists tend to assess, for instance, public opinion, illustrating pitfalls — both unintended and otherwise — with engaging examples. Baker, in contrast, devotes his longer and denser book to exposing the fragilities of science across its entire spectrum. His slant is set out on the first page of his acknowledgements: “Shortcomings in the scientific process itself and the nepotism that masquerades as the peer review of both manuscripts and grant applications conspire to push research programmes in favoured and fashionable directions irrespective of validity.”

Baker examines a series of current or recent disputes focusing on: sunscreens as a possible cause of cancer; cholesterol; clinical depression; bovine spongiform encephalopathy; global warming; conservation of biological diversity; and genetically modified (GM) foods. For each, he gives a comprehensive list of possible worries and reassurances, presented largely in a narrative style with very little analysis or synthesis.

For example, the chapter on sunscreens addresses concerns that chemicals applied to the skin as protection in fact promote cancers. I would have liked to have seen more discussion of possible differences between sunscreens that are based on simple physical barriers (such as the Australian cricketer Shane Warne’s colourful zinc cream), and more elaborate chemical creams. The extended discussion of GM foods underlines the differences in attitudes between North America and Europe, suggesting that the



How many? What was the true number of participants in the 1995 Million Man March?

answer may lie in differences in national character (“perhaps Europeans really are more easily panicked than Americans”). There is, however, no mention of the converse situation in the 1970s when, after the 1975 Asilomar meeting on the safety of recombinant DNA research, research on ‘gene-splicing’ went ahead in Europe, but met strong opposition in many North American university towns.

Baker concludes that “very often we can scarcely believe anything biological scientists tell us”. His final sentences are: “Today’s best hypothesis once properly tested may well be tomorrow’s joke, even if it’s not so funny. But eventually, the correct hypothesis will come along and stand the test of time, another robust pillar amidst the scientific colonnade of glass.”

I doubt the validity of this image of science. Baker’s relentless focus on what is uncertain, and on the “two sides” to every scientific story, seems to me fundamentally to mistake the nature of the scientific enterprise. Behind us lies a solid architecture of columns, the well-understood body of science. At the frontiers, where the vexatious problems lie, we do not have glass columns mistaken for the real thing, but rather workers assembling the masonry for what will be tomorrow’s columns, and arguing, indeed

often quarrelling, as they do so. Most of them recognize that, at the edges of the known, what we have is sometimes “two sides”, but more usually it is a landscape of scientific hypotheses and opinions — for some issues single peaks, for others a complex topology. And, as time goes on and knowledge accumulates, these complicated and fuzzy landscapes become more sharply defined.

Best’s major theme is linked to Baker’s in recognizing the complexities and uncertainties that bedevil so many scientifically related public debates. His focus, however, is on statistical studies and their presentation. Most earlier books in this area, such as Darrell Huff’s classic *How To Lie With Statistics* (Norton, 1954), deal mainly with the technical apparatus of statistics and its abuse, especially in misleading visual displays. Best, I think, goes deeper, explaining how inherent uncertainties and imprecisions in the area of human social behaviour can affect our ability to gather and interpret statistical information about ourselves.

His book is organized around a few central themes, among them the basic causes of misleading statistics (bad guesses, deceptive definitions, confusing questions and biased samples), “mutant statistics”

AP/CHARLES PEREIRA/US PARK SERVICE

(how good statistics can be misunderstood or misrepresented, knowingly or otherwise) and problems in comparisons (across time, space, groups). The narrative flows easily, and all the points are driven home with engaging examples from real life.

The book starts by examining the opening line of a recent PhD proposal: "every year since 1950, the number of American children gunned down has doubled." Starting with one death in 1950, this implies that more than the current world population would have been dead by the early 1980s. Best tracks the reference down to a colourful mutation of a 1994 report that "the number of American children killed each year by guns has doubled since 1950".

I often cite the well-known statistic that two per cent of Americans believe they have been abducted by aliens and returned to Earth. Sadly, Best explains the source: the researchers thought a direct question about abduction might be off-putting, so instead they devised five indicative symptoms. One was: "Have you experienced waking up paralyzed with a sense of a strange person or presence or something else in the room?" Scoring four out of five positive responses apparently equated to affirming abduction.

In a more constructive vein, Best shows how we can test for racial bias in police arrests. Suppose we find that among 100 white and 100 black youths, 10 and 17, respectively, have experienced arrest. This may look plainly discriminatory. But suppose we then find that of the 80 middle-class white youths 4 have been arrested, and of the 50 middle-class black youths 2 arrested, whereas the corresponding numbers of lower-class white and black youths arrested are, respectively, 6 of 20 and 15 of 50. These arrest rates correspond to 5 per 100 for white and 4 per 100 for black middle-class youths, and 30 per 100 for both white and black lower-class youths. Now, better analysed, the data suggest effects of social class, not race as such. I also especially liked Best's lucid and comprehensive analysis of the heated dispute between the Muslim leader Louis Farrakhan and the Washington Park police over how many actually marched in the Nation of Islam Million Man March in 1995.

In summary, both books address contemporary issues of science in society. I found Best's book a delight. Always engaging, it is accessible to a lay reader, yet will reward the expert; the examples it gives could enrich both a primary schoolroom and a university lecture hall. Baker's book is, according to your taste, comprehensive or long-winded, but it contains a great deal of information and opinion, and some readers will relish the polemics. ■

Robert M. May is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.



Jewels on the wing

Unique to the Americas, the 300 or so species of hummingbird range in their habitats from the snow line of the Andes down to the lowland rainforest and coastal mangrove swamps. The long-tailed hermit (*Phaethornis superciliosus*), shown here visiting a *Heliconia* flower, is just 15

centimetres in length and can consume amounts of nectar equivalent to its own body weight. *The World of the Hummingbird* by Robert Burton (Firefly, \$40, £27.95) tracks the natural history and lifestyle of the diverse species of these "flying jewels", with their unique flight mechanism.

Science and war

Le scientifique et le guerrier

by Jean-Jacques Salomon

Éditions Belin: 2001. 160 pp. 12.2 euros

John Ziman

How is it that such a good thing as science is so closely involved with such an evil as war? Recent horrifying events might seem to make this faustian question particularly 'timely', but surely no more so than for the past 50 years. Why have there been so few systematic attempts to answer this question? With robust eloquence, Jean-Jacques Salomon at last slashes open the shroud of our denial. Here is a book that should be taken to heart by all who take pride in our calling.

There is no denying the evil of war. But its rank against other evils has long been debated. The concept of a 'just war' has an honourable tradition. Some ways of killing people have always been considered peculiarly 'dirty' — especially when they were first invented. As a prime source of military innovation, science is thus deeply implicated. But as Salomon points out, such debates invoke ethical concerns that are impossible

to consider in 'scientific' terms. Scientists developing new weapons often claim that they are not competent to evaluate 'non-technical' considerations. So do mercenaries and prostitutes.

The innate goodness of science cannot be demonstrated 'scientifically'. Indeed, one of the firmest beliefs of scientists is that the knowledge they produce has no moral attributes. Not unreasonably, they refuse to take responsibility for the way their findings are ultimately used. So they have to balance the manifest benefits of science — for example, in medicine — against its equally manifest disbenefits — for example, the intensification of war. The empirical test is unconvincing.

The last resort is an overriding faith in the virtue of 'the truth'. Unfortunately, this is a metaphysical principle that cuts no ice in practice. So 'the truth' is transformed into a holy aura surrounding the most distinguished scientific truth-finders — think of Albert Einstein. It is not obvious, however, that his sincere, strongly avowed personal pacifism flavours the entire scientific endeavour. Salomon rightly scorns the 'schizophrenia' of other scientists, such as the eminent physicist Freeman Dyson, who are just as committed to seeking fundamental truths and yet participate actively in military research. I would add

that their subordination to such devious interests seriously compromises their credibility as 'public' scientists.

Whatever its moral attributes, science requires a strong rational intelligence. So scientists imagine that science could be used to build a 'planet without frontiers', modelled perhaps on the scientific community. Communists, for example, believed theirs was a 'scientific' political system that would produce a frontierless world. Well, that conceit collapsed unobtrusively during the cold war. More modestly, scientists hold that the 'scientific attitude' fits them uniquely for peacemaking. Certainly, the Pugwash movement played an important mediating role during the cold war. But as Salomon makes clear, the scientists on both sides who took part in these unofficial discussions about nuclear arms control were always kept on a leash by their political masters. This period may also have been unusual, in that rational dialogue was made possible by the 'part-time' engagement of many of them in nuclear-weapon development and their shared understanding of the technical parameters of deterrence. Again, the apparently 'scientific' problem of detecting violations of the nuclear-test-ban treaty was really hedged around with 'non-scientific' considerations of political acceptability. Anyway, these conditions no longer hold, especially for states where scientific institutions and élites carry little political weight.

The universality of science brings scientists together on an international level. They thus have unusual opportunities to initiate confidential negotiations, or publicly exemplify improved relationships, between competing nation states. Cross-border institutions such as CERN, the European particle physics laboratory near Geneva, pioneered the path to the economic and political unification of Europe. But the virtual 'internationale of the savants' — the republic of learning — never seriously competes with real national allegiances. Nor does it seem that scientists are uniquely qualified to perform quasi-diplomatic roles in preventing war: in present-day conditions, theologians, bankers or lawyers may be just as effective.

In sum, the faustian question falls apart. Science is not peculiarly good in any of the ways that war is peculiarly evil. Salomon doubts whether the scientific community is any more homogeneous or unified on such matters than any other profession. Most scientists find nothing wrong with the fact that a considerable proportion of them — Salomon doesn't give a figure, but my guess is about 20% — are now linked directly or indirectly to the military-industrial complex. Only a small minority peer over the edges of their specialist niches at the political ecology of their work. So the promotion of peace by organizations such as Pugwash is not some sort of atonement for our nuclear sins. It

bears witness to the fading ideal, "if not of a community, at least of individuals who care for the ends served by works of the mind". ■

John Ziman is at 27 Little London Green, Oakley, Aylesbury HP18 9QL, UK.

The American dream personified?

Smithsonian Institution Secretary,
Charles Doolittle Walcott

by Ellis Leon Yochelson

Kent State University Press: 2001. 832 pp. \$55

Richard A. Fortey

Charles Doolittle Walcott had his fingers in so many pies it is astonishing that he was able to find any available with which to write his monographs. At about the time of the First World War he managed to be secretary of the Smithsonian Institution in Washington, president of the National Academy of Sciences, chairman of the National Aviation Board, and Carnegie Institution of Washington regent, as well as being on intimate terms with all the politicians of his day — to say nothing of trying to set up a national art collection in the United States and guiding government policy in matters of aviation and forestry. I have probably forgotten one or two others.

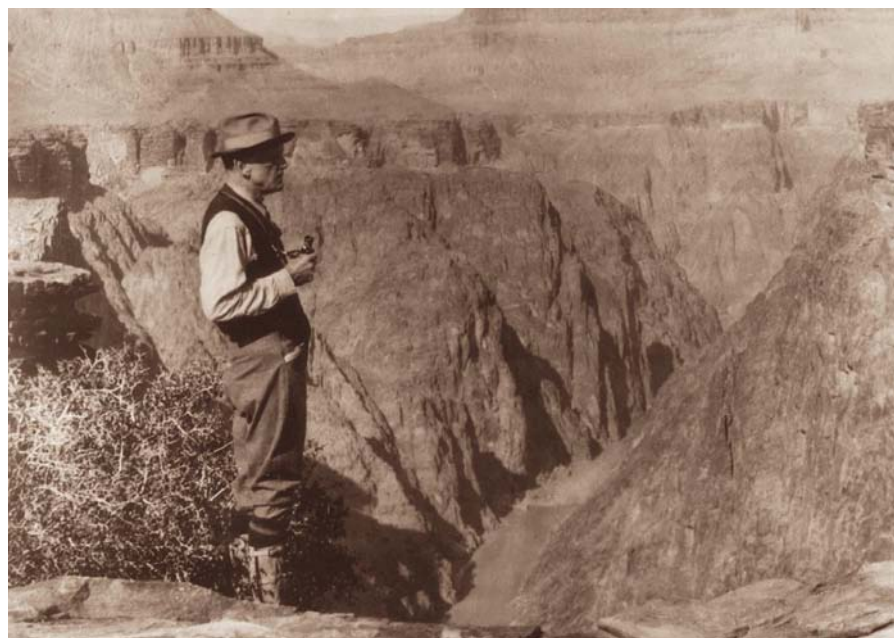
Walcott also managed to churn out huge volumes of basic science on the Cambrian rocks of North America. He is probably best known today for his discovery of the Burgess Shale — that remarkable assemblage of 'soft-bodied' fossils that allows us unprecedented insight into the marine life of the Cambrian. To describe Walcott as energetic is inadequate. Those of us who try to square the

necessities of administration with the demands of research can only feel abashed. Mind you, we do tend to take Christmas Day as a holiday.

Much of Walcott's research is still regarded as fundamental to geology today, and most of it was carried out with the aid of packhorses and pemmican. And if you add that during this period of his life he tragically lost a wife and two sons, you begin to wonder of what kind of steel he was constructed.

Walcott's is an important life, not least in the history of civil science, and it is high time it was documented. The first volume of Ellis Yochelson's account took us through the 'rags-to-riches' story of Walcott's early days: he was a self-educated farm boy, and came up the hard way through diligence and intelligence. The second volume reviewed here sees him as an established, not to say Establishment, scientist. By the end of his life he had so many honours that he scarcely bothered to record a new one in his diary. In some ways, the early days were more intriguing. He had to negotiate some particularly difficult characters — for example, the father of North American palaeontology, James Hall — on the way up. But in later life he had the ear of presidents to smooth out difficulties (can you imagine today's director of the British Museum making a New Year's Day call on the UK prime minister, Tony Blair?). Yochelson's account is almost bruisingly detailed — we find out what Walcott did on virtually every day of his life, stopping only at the lavatory door.

Many of those days followed a closely similar mould: a.m., called on important people and did some fixing on executive committee; p.m., found time for research; evening, made speech. As an account of a compulsively busy man it has a fascination



Moment of repose: but Walcott, best known today for discovering the Burgess Shale, drove himself hard.

all of its own, and it is probably necessary to set out these facts before some more critical account of Walcott's role can subsequently be written. Walcott achieved so much because he was so — to use the horrid jargon — compartmentalized. His life followed a productive routine. Every year he did his summer fieldwork in the Canadian Rockies, and it has to be said that one year's fieldwork reads a tad like any other, just as his days had a certain sameness. If genius be truly 99% hard work, then this man was undoubtedly a genius.

According to Yochelson, Walcott had one real brush with controversy (he won virtually all his scientific affrays) and it was nothing to do with the Cambrian. He offended the surviving Wright brother by claiming that his predecessor at the Smithsonian, Samuel Langley, was the true pioneer of manned flight. So offended was Orville Wright that he presented the original flying machine to the Science Museum in London rather than keep it in his native land (many years later, it was returned). It is odd that Walcott's diaries treat this business with such sangfroid — it seems that, compared with getting on with his fieldwork, even a scandal was of little importance.

Yochelson tends to settle all accounts in Walcott's favour, and the reader gets the impression that his admiration for this prototype Washington mandarin borders on idolization. As an embodiment of the American dream, Charles D. Walcott can scarcely be beaten, and Yochelson evidently despises the current obsession among biographers for finding feet of clay. But a more critical eye might have revealed greater insight into the man behind the endless memoranda and monumental monographs. ■

Richard A. Fortey is in the Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Elementary tales

Nature's Building Blocks: An A–Z of the Elements

by John Emsley

Oxford University Press: 2001. 552 pp. £20

Andrea Sella

Together with the atom, the elements form the other fundamental paradigm of chemistry. Although the term itself originated with the Greeks, the modern meaning of the word 'element' has gelled only over the past two centuries through the often mind-numbing struggles of pioneering chemists. For the past half-century the picture has been fairly static, apart from the few stragglers at the bottom of the periodic table

ELEMENTS					
	Hydrogen	1		Strontian	46
	Azote	5		Barytes	68
	Carbon	5		Iron	50
	Oxygen	7		Zinc	56
	Phosphorus	9		Copper	56
	Sulphur	13		Lead	90
	Magnesia	20		Silver	100
	Lime	24		Gold	190
	Soda	28		Platina	190
	Potash	42		Mercury	167

Foundations of chemistry: John Dalton's table of the elements, produced in 1805.

(pace Glenn Seaborg, discoverer of ten of them) that Tom Lehrer referred to as "those that haven't been discavard". Although new elements don't turn up that often these days, the hundred or so known elements are so varied in their behaviour, history and applications that it was only a matter of time before someone got stuck in to tell their tales.

Nature's Building Blocks, a layman's field guide to the periodic table, is a direct outgrowth of and companion volume to John Emsley's popular (among chemists, anyway) compendium of numerical data, *The Elements* (Oxford University Press, 1998). Emsley adopts the discursive approach that was so successful in his *Molecules at an Exhibition* (Oxford University Press, 1999), and gleefully describes the inhabitants of this world.

What for many might be a dry and dusty collection of facts has been turned into an amusing and finely crafted set of mini-biographies. Each element is introduced by a brief history of its discovery, together with a listing of its names in a number of languages (but why are they mostly European, especially when the Italian, Spanish and Portuguese are often identical? Would not Russian, Japanese or Chinese have been more widely useful?). For certain 'old' elements such as sulphur, useful Sanskrit, Hungarian and even Maori names are provided as well. The key physical properties of the element are seldom the first to be listed. Instead, they jockey for pole position with other issues more relevant to real life. Sixteen elements, for example, are singled out as "elements of war".

For the true magpie, however, the most amusing sections of the book are the text boxes and the little parting shots at the end

of each chapter, the "Elements of surprise". These sections gather together a wild variety of otherwise unclassifiable stories associated with each element. Thus, we learn that indium shrieks when bent, and that magnesium will not ignite on welding. The entry for neon discusses the suggestion that J. Norman Collie, the chemist and mountaineer who succeeded William Ramsay as professor of chemistry at University College London in 1913, might have been a model on whom Arthur Conan Doyle based the character of Sherlock Holmes. For thallium, a favourite of crime writers, the conviction of Graham Young — through forensic analysis of the cremated remains of one of his victims — provides a cautionary tale. Those whose interests lie in the more lubricious corners of history might have a look at "Mercury in famous men", which exonerates several individuals unjustly accused of having contracted syphilis. Thus, the elements come alive through stories and anecdotes. Emsley has cast his net wide and drawn on a huge range of material — this is a book for browsing almost at random.

Teachers of chemistry will find the book a marvellous source of titbits that can be used to leaven potentially dry lecture material. But the more general reader will find plenty to be intrigued and amused by. It is odd, however, that the book is almost entirely devoid of illustrations, apart from examples of different layouts of the periodic table (at the very end), and the elegant but repetitive picture of an atom which forms the heading to each chapter. It might have been more helpful to use a small schematic of the periodic table to make it easier to see the connections and relationships between elements. There is also no index, making it tricky to find half-forgotten items.

If one were being uncharitable, one might also point out a few small errors (the lanthanides were discovered in Vaxholm, Sweden, not Vauxholm, which sounds suspiciously like a railway station in south London). And there is the odd howler (is sunlight really yellow because of emission from sodium atoms? Has Emsley forgotten about black-body emission?). Also, some omissions have, no doubt, resulted from an attempt to keep the book to a sensible length.

But these are little more than quibbles. This is a fine, amusing and quirky book that will sit as comfortably on an academic's bookshelf as beside the loo, to be browsed and savoured in idle moments. Can it be long before we see a series on television called "Meetings with remarkable elements"? ■

Andrea Sella is in the Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.

Nature's helping hand

W. F. Bynum

Six centuries ago, the Italian humanist and poet Petrarch remarked that if a hundred, or a thousand, people of the same constitution and illness were divided into two groups, one under medical care and one entrusted to nature, the latter would show more cases of recovery. For many years, popular wisdom held that "there is much difference between a good and a bad physician, but not any between a good physician and none at all". In fact, L. J. Henderson, the famous Harvard biochemist, maintained that it was not until around 1910 that "a random patient, with a random disease, consulting a doctor chosen at random had, for the first time in the history of mankind, a better than fifty-fifty chance of profiting from the encounter".

Such observations invite a range of responses. A cynic might remark that Henderson's favourable odds have yet to be achieved even today, but a more benevolent view of medicine's past would attribute much to sensible advice, pastoral care, a few good remedies and the placebo effect. Whatever one's view, however, the puzzle remains of how medicine survived as a social institution despite centuries of therapeutic impotence, incompetence and pernicious meddling.

The Hippocratic physicians identified one potential answer: the healing power of nature. Doctors, they taught, are merely nature's servants. They took their diagnostic and therapeutic cues from what they could observe at the bedside — sick people, especially patients suffering from acute illnesses, often sweat, vomit, have diarrhoea, are pale, flushed or jaundiced, cough up phlegm or blood, lose their appetites, and develop pustules or

rashes. The Hippocratics interpreted these signs and symptoms as evidence that the body is a marvellous mechanism with an innate capacity to restore the natural humoral balance that constitutes health. Their ministrations were generally aimed at assisting and encouraging these natural processes.

This doctrine of the *vis medicatrix naturae* provided a rationale for much of the Hippocratics' activity. Inflamed limbs and fevered bodies turn red, clear evidence of too much blood. Why not therefore remove some through phlebotomy? Similarly, cathartics, emetics, sudorifics, diuretics and errhines were given to stimulate purging, vomiting, sweating, urination or nasal discharge, all frequent phenomena at the sickbed and presumably, therefore, part of the body's natural defences. To the Hippocratics, the art of medicine consisted of knowing when, and how much, to help — as well as when to stop.

Hippocratic humoralism dominated Western medical thinking until the eighteenth century. The doctrine of nature's healing power enjoyed reasonable assent, although its theoretical basis was reconceptualized by several doctors. In the seventeenth century, Jean Baptiste van Helmont saw healing as a manifestation of the workings of the archæus, a vital principle that he located in the stomach. Johann Wepfer visualized it as a function of the 'president' of the nervous system, whereas Georg Stall believed that it was diffused throughout the whole body.

Running through much of this debate was the anxious suspicion that the doctrine rather minimized the role of the expert in the treatment of illness within an unregulated medical market-place. If nature knows best, why bother to consult a doctor? A typical

Healing

The Hippocratic physicians identified the healing power of nature.

Doctors, they taught, are merely nature's servants.

attitude was that of John Conolly, who in 1859 acknowledged the power of natural healing, yet concluded that neither doctors nor their patients should ever "trust entirely to the *vis medicatrix naturae*, except in the case of the most trifling injuries, where a process of very small extent is all that is required for the cure".

By Conolly's time, both the authority of Hippocrates and the doctrine of *vis medicatrix naturae* were being appropriated by several sectarian challenges to what Samuel Hahnemann, the founder of homeopathy, had called the allopathy (treatment by opposites) of traditional Western medicine. Most nineteenth-century medical sectarians — practitioners of homeopathy, chiropractic, hydrotherapy, osteopathy and naturopathy — called upon Hippocrates, the 'father of Western medicine', to sanction their endeavours. The originator of modern naturopathy, Benedict Lust (his modern followers are at pains to point out that his name is pronounced 'Loost'), was inspired by "natural therapies used successfully since ancient times" to create his own philosophy of health and the treatment of disease. Search the Internet for *vis medicatrix naturae* and you will find yourself in the land of what we now politely call 'alternative' or 'complementary' medicine.

Hippocrates has thus become the patron saint of virtually all occidental medicine, scientific or otherwise. This status rests squarely on the Hippocratic insight dubbed "the wisdom of the body" by W. B. Cannon. Readers of *Nature* may well seek the source of that wisdom in the biological mechanisms of evolutionary adaptations. Indeed, the modern scientific custodians of the Hippocratic doctrine are the advocates of darwinian medicine, who question whether it is wise to provide symptomatic relief for the multitude of self-limited illnesses to which our flesh is heir. ■

W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 24 Eversholt Street, London NW1 1AD, UK.

FURTHER READING

Neuburger, M. *The Doctrine of the Healing Power of Nature Throughout the Course of Time* (New York Homeopath. Coll., New York, 1932).

Warner, J. H. *The Therapeutic Perspective: Medical Practice, Knowledge, and Identity in America, 1820–1885* (Harvard Univ. Press, Cambridge, Massachusetts, 1986).

Nesse, R. M. & Williams, G. C. *Evolution and Healing* (Weidenfeld & Nicolson, London, 1995).

They also serve who merely sit and wait: doctors may help, but nature often effects a recovery by itself.



See potassium run

Christopher Miller

Nearly all cells have membranes spanned by potassium-conducting channel proteins, without which your nerves (and much else) simply wouldn't work. Ion permeation through these channels can now be seen in dazzling detail.

Scientific discovery has its breakthroughs, and breakthroughs their aftermaths. Three years ago, Roderick MacKinnon's group at Rockefeller University in New York achieved a stunning advance in molecular neurobiology with the first atomic-resolution picture of a potassium (K^+) channel¹. Three papers in this issue²⁻⁴ now give us a mechanistic and dynamic view of how these pore proteins work.

Potassium channels are important membrane-spanning proteins that directly catalyse the ionic movements required to generate and shape electrical signals in neurons (and also in many other cells not so well endowed with brain-associated pizzazz). The original structure¹, though a bit fuzzy at 3.2 Å resolution, evoked gasps as it revealed the molecular design of these pores. It gave hope for an eventual understanding of the perplexing functional hallmark of K^+ channels — the combination of high selectivity and rapid throughput that allows K^+ to pass at such high rates that the protein seems to present no barrier at all, while simultaneously acting as a concrete wall to the smaller Na^+ ion.

The structure¹ sketched out the molecular basis of this specificity: a narrow 'selectivity filter' in the shape of an oxygen-lined electronegative tunnel in which dehydrated K^+ (but not Na^+) fits precisely. This structure rationalized why a K^+ ion is so willing to leave its thermodynamically comfortable home in aqueous solution to enter the pore in a largely dehydrated form; the channel interior mimics the embrace of the water molecules in the inner hydration shell surrounding the ion in solution.

And now, the aftermath. MacKinnon's group^{2,3} (pages 37 and 43) presents an improved structure of the K^+ channel, while Bernèche and Roux⁴ (page 73) carry out computer simulations of this same channel protein; together the results provide a deep understanding of how ions diffuse through this pore so rapidly. The new X-ray crystallographic structures, at a sharp 2 Å resolution, are dazzling. They offer three fundamental insights into the physical chemistry of ion conduction in biology: they show how ions are coordinated in the selectivity filter; they reveal the nature of K^+ ions in transit between the channel and the aqueous solvent; and they indicate the character of water in the cation's inner hydration shell.

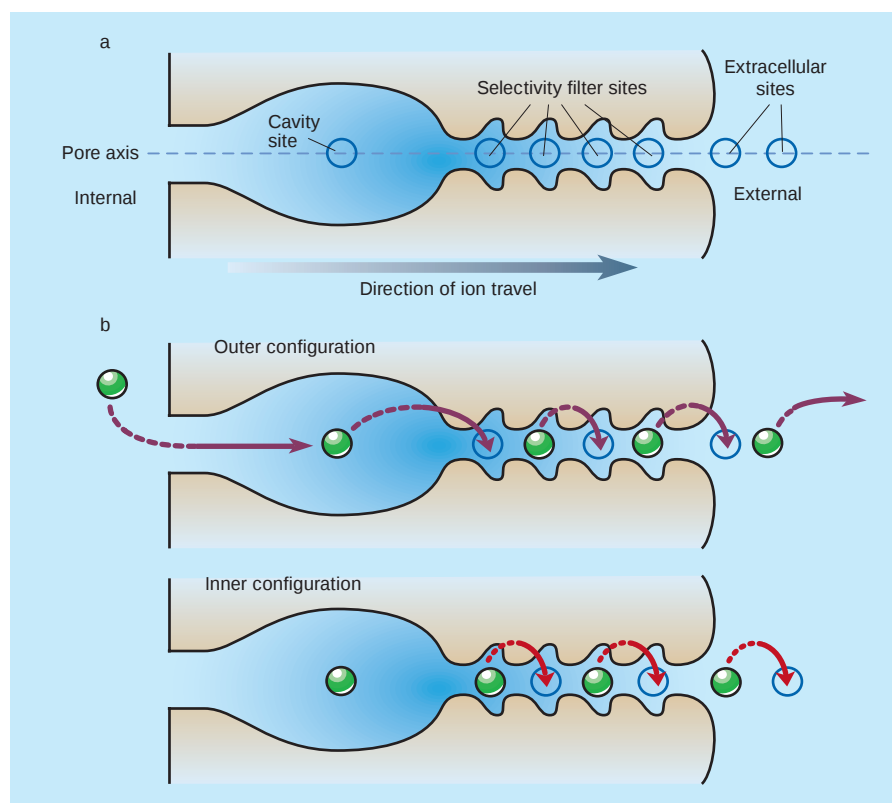


Figure 1 Permeation of K^+ ions through the pore of a K^+ channel, as surmised from the new results²⁻⁴ discussed here. **a**, There are seven main sites for ions along the pore axis: one in the pore cavity, four in the selectivity filter and two just beyond the external end of the pore. The cavity site is fully occupied, but (as indicated in **b**) only half of the remaining six are occupied at any one time. **b**, The two main ion configurations, known as outer and inner, that are postulated to exist within the pore. Purple arrows indicate ion shifts that are linked directly to concerted ion entry into and exit from the pore. Red arrows represent shifts within the pore without ion entry and exit. As shown here, then, ion passage through the selectivity filter and extracellular sites occurs in bucket-brigade fashion.

According to the familiar cautionary riff, X-ray crystallography is a snapshot-based technique impoverished in dynamic information about protein function. But these new pictures let the viewer imagine K^+ ions actually traversing the pore, because they catch the ions at several stages of movement. This is an unusual situation, unknown, for example, in crystallographic enzymology, which seeks to freeze the substrate at different stages of chemical transformation, each requiring a separate crystal structure. With K^+ channels, a single image makes these multiple stages visible all at once because, in contrast to enzymes, channels process several substrate ions simultaneously in bucket-brigade fashion^{5,6}. Potassium ions

are now seen in seven distinct sites along the pore-axis (Fig. 1a). Four of these reside in the narrow selectivity filter, and one in the wider hydrated cavity, as described earlier^{1,7}. By solving structures at varying ion concentrations, MacKinnon and colleagues argue that the four selectivity-filter sites are not all occupied simultaneously; rather, a pair of K^+ ions separated by a single water molecule shifts in a concerted fashion between two configurations within the filter — inner and outer — occupying each about half the time (Fig. 1b).

This ion-shifting between two pairs of positions, presumably on the conduction timescale of 10–100 nanoseconds, is at the heart of the permeation mechanism emerg-

ing from this work. As any long-distance cyclist knows, you make better time on level ground than in terrain with hills or valleys. It's the same for channel-dwelling K^+ ions, which permeate so rapidly because, as shown here by ion-occupancy measurements, the inner and outer ion configurations are precisely balanced in free energy, perfectly designed to eliminate barriers to high throughput. This point is buttressed by experiments with Rb^+ — a slightly larger, non-biological K^+ analogue indifferent to the demands of evolution — for which the two configurations are energetically out of balance; the resulting hilly landscape deduced from the X-ray data crisply explains the slower conduction of Rb^+ (refs 3, 8).

Two big surprises leap out of the results, and both enrich our comprehension of what life is like for ions diffusing through the channel. First, K^+ shows up at two additional on-axis positions just beyond the extracellular end of the pore, positions partially occupied by an ion shifting between the two. This ion is mainly in contact with aqueous solvent, so it is astonishing that it is localized enough to be visible in the electron-density maps used to pinpoint features by X-ray crystallography; strong electrostatic focusing by the channel surface is the likely explanation. Most dramatically, in the position closest to the pore entrance a K^+ ion is caught *in flagrante*, coordinated in front by four protein carbonyl groups reaching outwards, and behind by solvent; this must represent the long-postulated 'dehydration transition state' in which the ion sheds its water while entering the pore⁹. It is now seen not to be a high-energy transition-state at all, but rather a true intermediate, an integral part of the flat landscape.

Remarkably, in an independent computer simulation⁴, Bernèche and Roux anticipated that a K^+ ion would be localized in solvent, right at the two sites where it is actually seen in the X-ray work. This successful prediction in advance of the facts — a rarity in computational biochemistry — enhances the confidence of sceptical experimentalists in the methods and parameters used in this theoretical work. This tends to fortify an additional conclusion emerging from the work — that not only are the K^+ -binding configurations energetically similar, but so are the transitions between them. This means that the entire conduction process is energetically barrierless. So, overall, K^+ permeation looks rather like the concerted action of those steel pendulum toys known as Newton's balls. An ion entering the pore on the left expels an ion on the right into solution, while the four ions within and just outside the pore all shift one position to the right (Fig. 1b); if this occurs on energetically level ground, the process will be rapid.

The second surprise in the X-ray results is a structure that has often been imagined

but never before glimpsed: the complete inner hydration shell of an aqueous cation. The channel's central cavity has room for about 50 water molecules, along with the single hydrated K^+ ion suspended there. Here, eight of these waters are individually visible, with their oxygens packed directly against the K^+ ion at the corners of a twisted cube. This inner-shell geometry closely matches the arrangement of the K^+ -coordinating oxygens in the selectivity filter, and the authors wonder whether it might also mirror K^+ hydration in bulk liquid water. If so, they intimate, perhaps the four-fold architecture of K^+ channels evolved around the structure of hydrated K^+ to create an energetically welcoming home — the ultimate flat landscape — for ions forced to strip on their

swift journey from one side of the membrane to the other. ■

Christopher Miller is in the Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02468, USA.

e-mail: cmiller@brandeis.edu

1. Doyle, D. A. *et al.* *Science* **280**, 69–76 (1998).
2. Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. *Nature* **414**, 37–42 (2001).
3. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. *Nature* **414**, 43–48 (2001).
4. Bernèche, S. & Roux, B. *Nature* **414**, 73–77 (2001).
5. Hodgkin, A. L. & Keynes, R. D. *J. Physiol. (Lond.)* **128**, 61–88 (1955).
6. Neyton, J. & Miller, C. J. *Gen. Physiol.* **92**, 569–586 (1988).
7. Jiang, Y. & MacKinnon, R. *J. Gen. Physiol.* **115**, 269–272 (2000).
8. LeMasurier, M., Heginbotham, L. & Miller, C. J. *Gen. Physiol.* **118**, 303–314 (2001).
9. Hille, B. *Ion Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 2001).

Quantum optics

The atomic nanoscope

Andrew Steane

An ideal probe should be as small as possible so it doesn't interfere with the observation. When measuring the distribution of a light field, it seems that a single atom is up to the job.

The quest for the small has led to much ingenuity in fabricating tiny devices, and each step from micrometre to 100-nanometre to smaller scales is celebrated. One goal of such work is to reach the scale of the atom — the most sharply pointed needle one can readily conceive would have at its tip a narrow pyramid finishing with a single atom. But a sharp needle is still far from ideal as a probe: at the atomic scale it looks like a great bulky mass of material, and introduces a large disturbance into its three-dimensional environment. A radical alternative is to throw the body of the needle away altogether, and just keep the last atom. This probe would be a just a fraction of a nanometre in size in all directions — if only we could make use of it.

Guthöhrlein and co-workers¹ from the Max Planck Institute in Garching, Germany, have just taken such a step, by controlling and using a single atom to probe the structure of a standing wave of light. On page 49 of this issue they reveal the most precise measurement yet of the three-dimensional structure of a light-wave field. But this is only the first indication of what can be done with this system. The ability to control simultaneously both atomic and optical aspects of the system opens up a range of possibilities.

The Garching experiment combines the techniques of ion trapping and laser fluorescence, together with a high-quality optical cavity. The authors used a set of electrodes with oscillating voltages (known as a radio-frequency trap) to confine a single atom of

calcium that has a single positive charge. They then surrounded the trap with a pair of highly reflective concave mirrors to form an optical cavity — that is, a structure inside which light will reflect to and fro forever (except for small losses), just as sound does inside the body of a reverberating musical instrument such as a violin. In this situation the light forms a standing wave between the mirrors, so there is a series of regions of brightness and darkness, separated from each other by a quarter of a wavelength (Fig. 1). Because the wavelength in these experiments was 397 nm (this is just visible to the human eye as violet light), the light field inside the cavity goes from dark to bright to dark again every 198 nm.

The aim of the experiment was to measure this structure, and also the structure in the light field perpendicular to the standing wave, which consists of the Hermite–Gauss functions known to students of optics and quantum mechanics (see Fig. 2 on page 49). To do this, several features of ion-trap experiments come to the fore. First, ion traps are tight traps: even though the electrode structure providing the trapping is on the order of 1 mm in size, the ion is confined tightly enough to restrict its movements to a region a few tens of nanometres across. This is like confining a charged coin in a region the size of a wine glass by using charged metal plates situated a mile away.

Second, to achieve this sort of confinement the ion also needs to be cold, because otherwise its natural thermal vibrations would make it explore a larger region inside

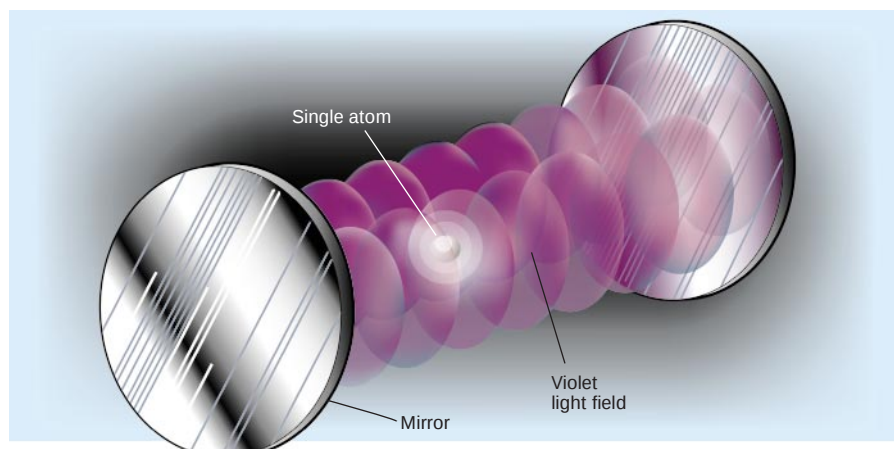


Figure 1 The amazing atom probe. In the experiment of Guthöhrlein *et al.*¹ a single atom of calcium is held suspended in space by electric fields. Two mirrors enclose the trapped atom, forming an optical cavity, within which a light field forms a standing wave. The atom fluoresces in the violet light field bathing it. As the atom moves across the light field, the scattered fluorescence is collected and its brightness as a function of the atom's position is used to construct a three-dimensional map of the light field (shown at reduced size for clarity). At the scale of the diagram, the mirrors and electrodes of the trap would be approximately 10 metres away from the atom. The exquisite control over light and matter required for a single-atom probe could find many other applications — from single-atom lasers to quantum computing.

the trap than desired. Fortunately, laser cooling is easily applied to trapped ions, and temperatures of less than a millikelvin are not uncommon. Finally, atoms shine very brightly for their size: a single atom of calcium, for example, will scatter as much light from a laser beam with a wavelength of 397 nm as a ground glass sphere of 600 times larger radius would. This is because of the phenomenon of resonance — the merest glimmer of incident radiation will cause the atom to 'sing' at its resonant frequency.

The hurdle faced by Guthöhrlein *et al.*¹ was to combine a high-quality optical cavity with the ion trap. This is a difficult undertaking because the cavity mirrors, if they are to be highly reflecting, cannot be made of a conducting material, so they are liable to become electrostatically charged and disturb the proper functioning of the trap. Also, when loading atoms into the trap there is a risk that a very thin coating of the trapped atoms will degrade the mirrors. A few laboratories have recently begun to tackle these problems. One feature of the Garching work is the use of two traps with a link between them. The ion is loaded into the first trap 25 mm away from the mirrors, and is then moved into the region between the mirrors by switching the trap voltages.

With the atoms in place, all that remains is for the experimenters to move the ion around and detect the light it emits with a brightness proportional to the intensity of the standing-wave light field bathing it. To reach their reported resolution of 60 nm this movement has to be highly controlled, and in the experiment it proved easier to leave the ion trap fixed relative to the rest of the laboratory, and then move the light

field by controlling the position of the cavity mirrors with piezo-electric transducers. In this way, they obtain beautifully clear three-dimensional maps of the light field.

Cancer

How cells choose to die

David Lane

Humans have several ways of keeping cancer at bay, and the p53 protein forms a crucial part of this self-defence. The newly discovered action of a p53-binding protein helps explain how cells respond to p53.

As their name suggests, tumour-suppressor genes encode proteins that help to prevent cancer. They generally do this by keeping the number of cells in multicellular organisms under strict control. The p53 protein is one of the best-known tumour suppressors¹, and its ability to keep cell numbers down seems to be essential in most animals, from worms² to fruitflies to humans (although not all of these species suffer from cancer). In vertebrates, dividing cells can respond to p53 in three main ways: they can stop dividing, become specialized or commit suicide. But which of these processes is needed for p53's ability to block tumour development, and how do cells choose the appropriate response?

An exciting paper by Samuels-Lev and colleagues³, published in *Molecular Cell*, suggests some new answers. The authors show that a long-known partner of p53 — a protein called p53BP2 — is actually a fragment of a larger protein, which they dub ASPP2. This protein, like its close relative ASPP1, steers cells towards death by binding

But the main practical implication of this experiment is not the creation of a new type of microscope, or even nanoscope. For the first time it offers the combination of an almost fully controlled atom and an almost fully controlled light field. This is a powerful combination, because it is a fundamental system in quantum optics and quantum computing, and brings with it the prospect of complete coherent control of a light-and-matter system². Coherence is a key feature of quantum systems — roughly speaking it means that all the components are in step. Light is the most natural conveyor of quantum information, and controllable matter is the most natural processor of quantum information. Although considerable effort is now going into building controllable quantum systems at the nanometre scale for quantum computing applications, it may be that nature has already provided the ideal processing elements in the form of single, isolated atoms.

Andrew Steane is at the Centre for Quantum Computing, Clarendon Laboratory, Parks Road, University of Oxford, Oxford OX1 3PU, UK.
e-mail: a.steane1@physics.ox.ac.uk

1. Guthöhrlein, G. R., Keller, M., Hayasaka, H., Lange, W. & Walther, H. *Nature* **414**, 49–51 (2001).

2. Eschner, J., Raab, Ch., Schmidt-Kaler, F. & Blatt, R. *Nature* **413**, 495–498 (2001).

to p53 and enhancing p53's capacity to switch on death-promoting genes.

Like the story of p53 itself, the history of the ASPP proteins is convoluted. The p53BP2 fragment was identified⁴ by Stanley Fields in one of the first 'two-hybrid' screens (a technique used to identify protein–protein interactions). Then, with remarkable speed, the crystal structure of a complex of the core region of p53 and the carboxy-terminal part of p53BP2 was determined⁵. The structure revealed how the two proteins fit together, and provided some interesting details. For example, the amino acids on p53 involved in binding p53BP2 are often mutated in human tumours, and p53's DNA-binding and p53BP2-binding domains overlap.

But the researchers were unable to explain what p53BP2 does, creating an unusual situation in which a protein's structure was understood before its function. Since then, a dribble of papers has detailed the interactions of p53BP2 with other molecules, as well as the fact that its expression is altered in human cancers. None of this was

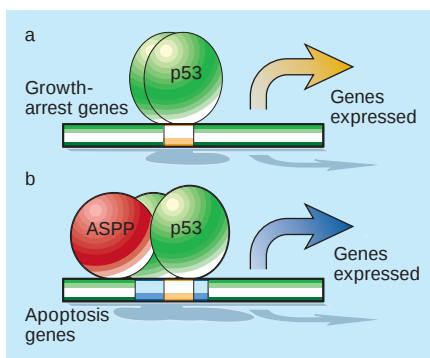


Figure 1 To die or stop dividing? Cells make this decision in response to activation of the p53 protein, which switches on target genes involved in either: a, growth arrest; or b, cell death (apoptosis). a, p53 normally recognizes sites in the control regions (promoters) of genes involved in growth arrest. b, As Samuels-Lev *et al.*³ show, however, when p53 forms a complex with the ASPP proteins, it interacts more strongly with sites in the promoters of apoptosis-inducing genes. The authors also analysed tumours from patients with breast cancer, and found that p53-induced apoptosis is lost either when ASPP proteins are no longer functional, or when p53 itself is mutated.

very compelling, however, and the importance of the p53–p53BP2 interaction itself was questioned by the observation that although p53 works inside the nucleus, most p53BP2 is found outside the nucleus⁶. The new study³ goes a long way to silencing these doubts and to assigning a vital role for p53BP2 in p53-induced cell death.

There are three broad ideas about how cells might choose between death (apoptosis) and simply halting their division (growth arrest) in response to p53. In the dose-response model, low levels of p53 induce growth arrest, and higher levels trigger apoptosis. By contrast, the 'cell-background' hypothesis argues that the response to p53 is decided by the cellular milieu. So, for example, cells expressing high levels of anti-apoptotic genes will be more likely to undergo growth arrest than apoptosis in response to p53. The final model proposes that the promoters (control regions) of different p53-responsive genes are recognized by differently modified forms of p53.

There is evidence to support all of these ideas. Cell lines containing certain p53 constructs undergo growth arrest when low doses of p53 are expressed, and apoptosis at higher levels⁷. In response to the activation of p53 by the drug leptomycin B, neuroblastoma cells undergo apoptosis whereas fibroblast cells — which contain high levels of anti-apoptotic proteins — simply stop dividing⁸. Finally, a mutant p53 with a subtle alteration in its DNA-binding specificity induces apoptosis but not cell-cycle arrest⁹, whereas other mutants — in which, interestingly, p53BP2-binding amino acids are

affected — have the opposite effect¹⁰. In many of these cases, p53-activated expression of the p21 protein seems to be crucial in inducing growth arrest and protecting the cell from apoptosis¹¹.

The new results support a variant of the third, promoter-specificity model. Samuels-Lev *et al.*³ started by clarifying the exact nature of p53BP2. Some sleuthing in the databases revealed two disturbing facts: that all previous studies had been carried out with a truncated form of the protein lacking the amino-terminal portion; and that p53BP2 has a very close relative. The newly named ASPP2 is the full-length p53BP2; ASPP1 is the relative. Both full-length ASPP proteins interact with p53, and, when artificially expressed in cells, dramatically increase p53's ability to induce apoptosis. Moreover, preventing the expression of the intrinsic ASPP1 and ASPP2 proteins partially inhibits the capacity of chemotherapeutic drugs to cause apoptosis, apparently by blocking p53's apoptosis-inducing ability.

So how do the ASPP proteins work? The authors show that the proteins cause p53 to bind specifically to, and activate transcription selectively from, p53-responsive pro-apoptotic genes (Fig. 1). In the same experiments, expression of the growth-arrest protein p21 was barely increased, if at all. In the authors' model, then, the ASPP proteins promote p53-dependent cell death by enhancing the activation of pro-apoptotic genes over growth-arrest genes. How might this be achieved? One possibility, supported by the new results, is that the whole ASPP–p53 complex acts at gene promoters, with the ASPPs subtly altering the ability of p53 to bind DNA.

It remains to be seen how important the ASPP proteins are in human cancer. As is so often the case, a rigorous genetic approach — using inactivation of both ASPP genes in

the presence and absence of p53 — will be needed to determine with certainty how big a role these players have. But tantalizing hints have already emerged from Samuels-Lev *et al.*'s quick survey of the loss of ASPP expression in breast cancers. The authors looked at normal and cancer tissues from 40 breast-cancer patients who expressed normal p53. In more than half of the patients, the expression of ASPP1 in tumours was lower than in unaffected tissue. Likewise, expression of ASPP2 was reduced in breast tumours in 9 of the 40 patients, and 8 of these also had low ASPP1 levels³.

These results raise the exciting possibility that, in losing functional ASPP proteins, tumour cells can block the p53-induced tumour-suppressor pathway. This theory will no doubt be put to the test by more studies of ASPP levels in tumours. Echoing the early days of p53 research, however, a word of caution is needed. It is possible that cancer cells might not always reduce the levels of ASPPs: they might instead express carboxy-terminal fragments of these proteins, which could prevent the full-length proteins from functioning and thereby promote cancer. A stringent and detailed analysis should, as always, pay dividends. ■

David Lane is in the Department of Surgery and Molecular Oncology, Ninewells Hospital Medical School, Dundee DD1 9SY, UK.

e-mail: d.p.lane@dundee.ac.uk

1. Vogelstein, B., Lane, D. & Levine, A. J. *Nature* **408**, 307–310 (2000).
2. Brent Derry, W., Putzke, A. P. & Rothman, J. H. *Science* **294**, 591–595 (2001).
3. Samuels-Lev, Y. *et al.* *Mol. Cell* **8**, 781–794 (2001).
4. Iwabuchi, K., Bartel, P. L., Li, B., Marraccino, R. & Fields, S. *Proc. Natl Acad. Sci. USA* **91**, 6098–6102 (1994).
5. Gorina, S. & Pavletich, N. P. *Science* **274**, 1001–1005 (1996).
6. Iwabuchi, K. *et al.* *J. Biol. Chem.* **273**, 26061–26068 (1998).
7. Chen, X. *et al.* *Genes Dev.* **10**, 2438–2451 (1996).
8. Smart, P. *et al.* *Oncogene* **18**, 7378–7386 (1999).
9. Saller, E. *et al.* *EMBO J.* **18**, 4424–4437 (1999).
10. Ludwig, R. T. *et al.* *Mol. Cell. Biol.* **16**, 4952–4960 (1996).
11. Polyak, K. *Genes Dev.* **10**, 1945–1952 (1996).

Surface science

The insulator uncovered

John B. Pethica and Russ Egdell

Microscopes that reveal the structures of surfaces one atom at a time are good at imaging conductors but perform poorly with insulators. That may be about to change.

The invention of the scanning tunnelling microscope (STM) sparked a revolution in atom imaging in the 1980s, and it has since become a mainstream tool. In the STM, a sharp metallic tip is scanned across a surface, causing a current of electrons to tunnel between the tip and surface. By keeping this current — and hence the gap between the STM tip and the surface — constant, the tip will naturally follow the contours of the atomic landscape and generate an image.

On conducting surfaces, such as metals and semiconductors, the STM can easily resolve single atoms, but insulators present a much bigger challenge.

A closely related technique that does not rely on electron tunnelling is atomic force microscopy (AFM), but this too has been defeated by most insulators. Now, on page 54 of this issue, Barth and Reichling¹ present atomic-resolution AFM images of alumina (aluminium oxide, Al₂O₃), a true insulator



100 YEARS AGO

Sir Hiram S. Maxim confirms the observation mentioned in our issue of October 17 (p. 607) of the attraction which certain sounds have for mosquitoes. Writing to the *Times*, he states that one of the electric lamps which he put up at Saratoga Springs, New York, in 1878, emitted a musical note; or rather the note proceeded from the box containing the dynamo machine under the lamp. One evening whilst examining the lamp he found that everything in the immediate vicinity was covered with small insects. They did not appear to be attempting to get into the globe, but into the box that was giving out the musical note. A close examination of these insects showed that they were all male mosquitoes... Sir Hiram Maxim remarks that "when the lamps were started in the beginning of the evening every male mosquito would at once turn in the direction of the lamp, and, as it were, face the music, and then fly off in the direction from which the sound proceeded... as the pitch of the note was almost identical with the buzzing of the female mosquito the male took the music to be the buzzing of the female." From *Nature* 31 October 1901.

50 YEARS AGO

There has been a lively discussion, in *Nature* and elsewhere, on the classification of recent fossil finds in South Africa. Prof. S. Zuckerman has rightly stressed the need for basing such decisions on statistical tests. But his findings have been at odds with those based on classical comparisons; so that anthropologists must now be in doubt whether their science has, indeed, anything to learn from statistics. This unhappy result can be traced to the piecemeal tests which have hitherto been used. A bone or a tooth is a unit; it is not a discrete assembly of independent measurements. To compare its single variates one by one, as Ashton and Zuckerman have done, is both inconclusive and misleading... The technique of multivariate analysis makes it possible to construct functions to discriminate at one time between a number of species, and we hope elsewhere to construct such functions for human, chimpanzee, gorilla and orang-utan teeth. These functions can, we believe, have an important place in biological classification, for which indeed they were first designed.

From *Nature* 3 November 1951.

that has long been considered the definitive test for AFM. The surfaces of insulators are in many ways as important as those of conductors, yet they remain largely unexplored — this work opens a new route to their investigation.

AFM works by measuring chemical, electrostatic or magnetic forces between the microscope tip and the sample. The tip is usually mounted at the end of a flexible cantilever whose oscillation frequency changes whenever the surface exerts a force on the tip. This versatile technique can be used to study processes ranging from biomolecule mechanics to the growth of thin films. Both STM and AFM are key tools for nanotechnology, and a casual observer might assume that they are well-understood 'commodity' items. So it is perhaps surprising that it has taken AFM so long to achieve atomic resolution of insulating surfaces. This was, after all, its original *raison d'être* and advantage over STM.

Following the invention of AFM it took ten years to produce the first atomic-resolution images of silicon². Moreover, STM can image most of the same surfaces as AFM, including moderate insulators, and has recently imaged surfaces of insulating diamond owing to an electron-injection technique³. Naturally, questions were asked about how atom-resolved AFM actually works: does it require some movement of electrons, or can it see atoms on a true insulator surface?

Alumina is one of the most important metal oxides, both as a traditional ceramic and in powder form as a support for dispersed metal catalysts. Yet despite a huge upsurge in interest in metal oxides, alumina surfaces have refused to yield their secrets. Techniques based on measuring various properties of electrons, such as diffraction, photoemission and STM, are difficult to apply to alumina because the bulk material is highly insulating. To circumvent these difficulties, scientists developed techniques to grow ultrathin and atomically ordered layers of alumina as 'models' for real alumina surfaces⁴. Electrons from the material below the thin film (usually a single crystal of NiAl) pass through the alumina film to the STM tip, but unfortunately these thin-film surfaces do not capture all the subtle complexities of 'true' alumina surfaces.

In particular, the basal surface of alumina crystals — the crystallographic (0001) face — can exist in several different phases, depending on how the crystal has been treated. Heating alumina at very high temperatures in a vacuum produces a so-called $\sqrt{31}$ phase, whose structure has been investigated by electron and X-ray diffraction techniques^{5,6}. This phase cannot be reproduced in experiments on thin alumina films. The original model for the $\sqrt{31}$ phase assumed that the surface atoms had a square arrange-

ment, but more recent studies using X-ray diffraction suggested a hexagonal structure⁶. The AFM images taken by Barth and Reichling¹ confirm that the surface atoms do indeed form hexagonal patterns. The authors also show how the surface structure changes when water is adsorbed.

How did they do it? Many groups with strong records in atomic-resolution AFM have tried to image alumina, but with no success. The key to Barth and Reichling's experiment is the careful preparation of a 'clean' surface and rigorous control of the tip, in particular minimizing energy dissipation during imaging. It is perhaps significant that their findings support a model of the (0001) face of alumina that has two layers of aluminium metal at the surface, which would mean the surface is metallic even though the bulk is insulating⁷. So surface conduction might still play a role in the AFM measurements. It remains to be seen if 'truly' insulating oxide surfaces, such as other crystal faces of alumina, can also be probed by AFM.

But we might also ask whether the concept of a 'true' insulator is appropriate when experiments involve just one or two atoms. Conduction occurs when electrons move from one place to another in a material and remain there when the original stimulus (voltage) to move is removed. Charges can also move in insulators, but they always return to their starting location when the voltage that moves them is reduced. Charge movement can happen over quite a long distance, depending on the 'insulating' material, and certainly further than the single-atom resolution scale of AFM and STM. The new charge location could therefore be some distance from the tip, and the resulting net movement of charge could explain the initially surprising successes of STM imaging of moderate insulators. The flow of charge can also be enhanced, for example by doping or heating, or by finding new mechanisms such as the electron-injection technique used for diamond³.

The AFM has been most productive⁸ when imaging surfaces that have some potential conducting ability or when electrostatic charges play a significant role in imaging. On insulating halides, such as CaF_2 , detailed modelling and experimental work have revealed that whether anions or cations are imaged depends on the electrostatic potential produced by the tip at the surface⁹. There may be a closer connection between AFM and STM when resolving atoms than was first expected. It is now known, for example, that significant local forces act during STM imaging, and the resulting changes in atom positions probably cause most of the image contrast in STM imaging of close-packed metals. It seems that forces and

electronic effects can influence both AFM and STM images.

Scanning microscopes will be vital elements in the atom engineer's toolbox, and our understanding of their operation continues to advance rapidly. We can look forward to finding a sound basis not only for imaging and spectroscopy of single atoms, but also their manipulation and the creation of wholly novel nanostructures. ■

John B. Pethica is in the Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK.

Russ Egdell is in the Inorganic Chemistry

Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK.

e-mails: john.pethica@materials.ox.ac.uk

russ.egdell@chem.ox.ac.uk

1. Barth, C. & Reichling, M. *Nature* **414**, 54–57 (2001).
2. Giessibl, F. J. *Science* **267**, 68–71 (1995).
3. Bobrov, K., Mayne, A. J. & Dujardin, G. *Nature* **413**, 616–619 (2001).
4. Libuda, J. *et al.* *Surf. Sci.* **318**, 61–73 (1994).
5. Chang, C. C. J. *Appl. Phys.* **39**, 5570–5573 (1968).
6. Renaud, G., Villette, B., Vilfan, I. & Bourret, A. *Phys. Rev. Lett.* **73**, 1825–1828 (1994).
7. Gillet, E. & Ealet, B. *Surf. Sci.* **273**, 427–436 (1992).
8. Yokoyama, K., Ochi, T., Sugawara, Y. & Morita, S. *Phys. Rev. Lett.* **83**, 5023–5026 (1999).
9. Foster, A. S., Barth, C., Schluger, A. L. & Reichling, M. *Phys. Rev. Lett.* **86**, 2373–2376 (2001).

Cell biology

Bacteria thread the needle

Craig L. Smith and Scott J. Hultgren

When bacteria attack another organism, one of the first steps is the injection of 'virulence effector proteins' into its cells. Two of the main players in such a system have been caught in action.

Many bacterial pathogens use a complex secretion system to inject proteins, known as 'effectors', into target host cells. This process is part of a molecular subversion tactic that allows bacteria to persist and cause disease, and requires specific molecular chaperones. Our understanding of how these chaperones work is boosted by a report on page 77 of this issue¹, where Stebbins and Galán describe a chaperone-effector complex from the type III protein secretion system used by many bacterial pathogens, including *Salmonella*.

Cellular life requires that linear polypeptide chains fold into functional three-dimensional domains, which are used to build up larger supramolecular structures. However, the exposure of hydrophobic surfaces during protein folding can lead to the non-productive aggregation of macromolecules. To combat this problem, cells use molecular chaperones, which, like human chaperones, have the job of preventing their charges (in this case proteins) from making unacceptable non-productive interactions, while encouraging them to encounter and interact with acceptable partners². To bring about the folding, assembly and secretion of effector proteins, pathogenic bacteria use specific chaperones, some of which function by donating missing steric information to couple folding³ with the simultaneous capping of interactive surfaces⁴.

Various bacterial pathogens produce a surface appendage known as the type III secretion apparatus. The type III structures (injectosomes) provide a means of squirting the effector proteins into cells; in turn, these proteins manipulate various cellular processes to the advantage of the bacterium⁵. Type III secretion appendages⁶ have a cylin-

dric structure like the needle of a syringe, and translocation of effector proteins through them requires specific chaperones⁷.

Stebbins and Galán¹ now present the crystal structure of a type III chaperone (SicP) bound to the amino-terminal domain of its effector protein (SptP). The SicP protein is a kidney-bean-shaped homodimer with much of its solvent-accessible surface covered with clusters of hydrophobic residues. These hydrophobic highways provide the interface where SptP binds. SptP

(residues 35–139) binds to SicP in an extended, non-globular conformation in which its secondary structures (α -helices and β -strands) remain intact (Fig. 1). The grooves and crevices in SicP that make intricate contact with the SptP fragment are probably the basis for type III chaperone specificity. The amphipathic helix in the chaperone was thought to be important in protein-protein interactions⁷, but it has no direct interaction with the effector and is not involved in chaperone dimerization. So it is likely to be a general structural feature of these chaperones.

The protein's structure was crystallized as a crossed-over SicP-SptP 4:2 complex (Fig. 1). Luo *et al.*⁸ have used light scattering to show that other type III chaperone-effector complexes exist in a 2:1 stoichiometry. In the crystal structure reported here, there is a potential domain swap involving helix 5. This domain swap may occur during the crystallization process and subsequently be stabilized by a disulphide bond that is observed in the crystal structure but not in the purified complex before crystallography. So, as the authors point out, it is possible that the biologically relevant complex has a 2:1 stoichiometry.

Type III effector proteins have one or more functional domains. For a protein to travel through the secretion needle, each domain must physically be able to fit inside the cylindrical cavity of the channel. If the domain is wider than the secretion tube, it can't pass through it. On the evidence of its previously solved crystal structure, SptP has two domains. The diameter of the secretion channel has been estimated to be about 30 Å

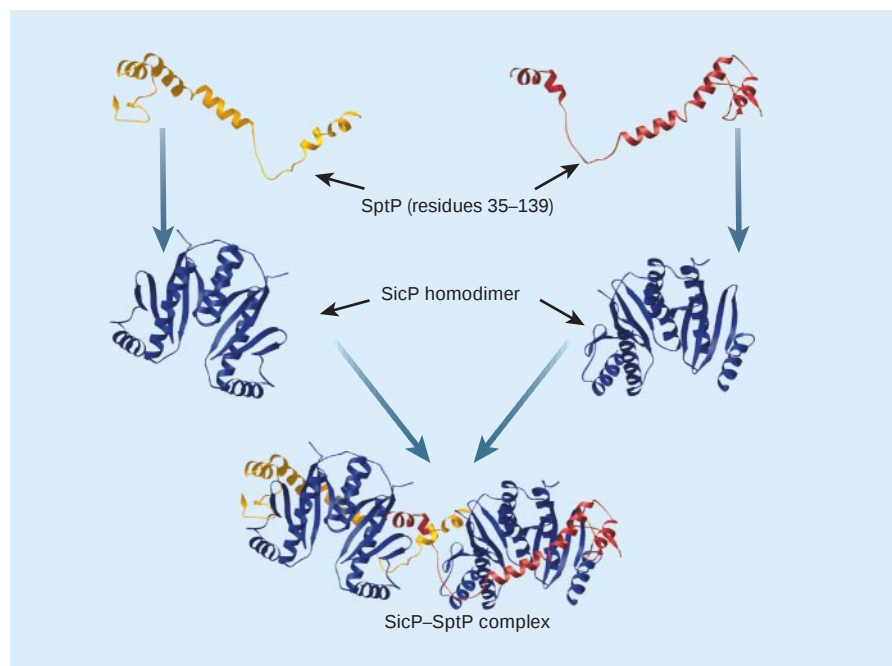


Figure 1 Formation of the complex between SicP (the chaperone) and SptP (the *Salmonella* effector protein). The non-globular, chaperone-binding domain of SptP (yellow and red) wraps around a SicP homodimer (blue). The crystal structure reported by Stebbins and Galán¹ is a 4:2 complex, created by the combination of two 2:1 SicP-SptP complexes.

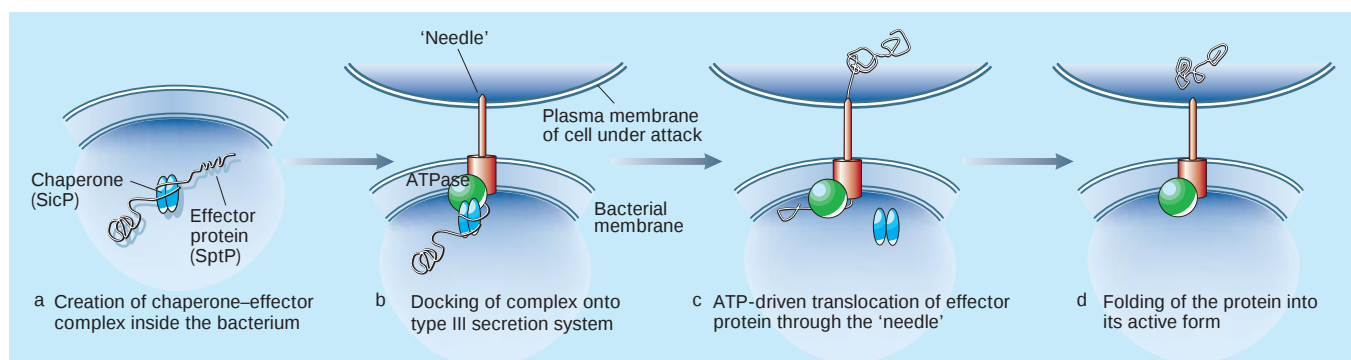


Figure 2 Model of chaperone-assisted type III secretion in *Salmonella*. a, Creation of the complex between the chaperone (SicP, blue) and the virulence effector protein (SptP, black) to be injected into the target cell. b, The complex docks onto the type III secretion apparatus (red), depicted here as lying between the bacterium's inner and outer membranes, with a 'needle' passing into the plasma membrane of

the cell under attack. c, Powered by ATP, the effector protein passes through the needle into the cell. d, The protein folds into its active form inside the cell, and subverts cellular function to the bacterium's advantage. The SicP–SptP structure produced by Stebbins and Galán¹ helps clarify events at the beginning of this sequence, but later events remain unclear.

(ref. 6). The dimensions of the SptP domains are approximately $74 \times 42 \times 39$ Å, which paradoxically seem too big to fit inside the secretion channel.

The structure presented by Stebbins and Galán explains at least part of this paradox. The SicP chaperone binds to the amino-terminal domain of SptP to keep it in a semi-unfolded conformation that would be able to fit into the secretion tube. But the other two domains not bound by the chaperone presumably exist in a folded state, and so would be too large to fit into the secretion tube. Indeed, Luo and colleagues' biochemical work⁸ on the chaperone–effector complexes they studied (SigE–SigD and CesT–Tir) suggest that the domains not bound to the chaperone remain folded. So either the channel has to get bigger or the domains have to unravel into a linear conformation.

Figure 2 provides an outline of how *Salmonella* injects effector protein into a cell that it is attacking. But we are still left with plenty of questions, especially as to how an effector protein is targeted to and travels through the injectosome and into the cell. The amino-terminal 35 amino acids were proteolytically cleaved before crystallization, suggesting that this region is also not protected by the chaperone. The region could be part of a motif that facilitates targeting of the complex to the type III secretion system, consistent with reports for other type III effectors in *Yersinia*^{5,9}. After targeting of the complex to the secretion tube, what is the mechanism of chaperone dissociation, and how are the folded domains of SptP unfolded so that it can move through the needle complex? Is ATP needed for dissociation of the SicP chaperone and/or to unfold the effector? How does SptP achieve its active conformation once it has been injected into the cell being attacked (in the absence of its chaperone)? Indeed, might effector proteins not even travel through the needle, but instead somehow pass into the host cell in an already folded state?

Stebbins and Galán's structure¹ is the first of its kind to be solved, and provides insight into type III secretion systems in general. It also reveals fundamental clues about chaperone function and mechanism of action, and will influence our understanding of the basic principles of protein folding and pathogenesis.

Craig L. Smith and Scott J. Hultgren are in the Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 South Euclid Avenue, St Louis, Missouri 63110, USA.

e-mails: smith@borcim.wustl.edu
hultgren@borcim.wustl.edu

1. Stebbins, C. E. & Galán, J. E. *Nature* **414**, 77–81 (2001).
2. Ellis, R. J. *Semin. Cell Dev. Biol.* **11**, 1–5 (2000).
3. Sauer, F. G. et al. *Science* **285**, 1058–1061 (1999).
4. Barnhart, M. M. et al. *Proc. Natl Acad. Sci. USA* **97**, 7709–7714 (2000).
5. Hueck, C. J. *Microbiol. Mol. Biol. Rev.* **62**, 379–433 (1998).
6. Kubori, T. et al. *Science* **280**, 602–605 (1998).
7. Wattiau, P., Woestyn, S. & Cornelis, G. R. *Mol. Microbiol.* **20**, 255–262 (1996).
8. Luo, Y. et al. *Nature Struct. Biol.* (2001); <http://dx.doi.org/10.1038/nsb717>.
9. Smith, C. L., Khandelwal, P., Kellikuli, K., Zuiderweg, E. R. P. & Saper, M. A. *Mol. Microbiol.* (in the press).

Theoretical physics

In search of exact solutions

Michel Héritier

Many-body systems, such as electrons flowing in a superconductor, are among the most difficult theoretical problems to study. A new family of exactly solvable models may offer some answers.

During the past 40 years, understanding quantum systems involving many particles has been one of the main goals of theoretical physics. This kind of research is designed not to work out the laws of interactions between particles, but, assuming that these interactions are already known, to calculate their effects on a system of n particles. A system where n is large constitutes the 'many-body problem', and can exhibit rich behaviour, including phase transitions, superconductivity and Bose–Einstein condensation. The impact of exactly solvable theoretical models on research into these systems is undeniable. With few exceptions, previous exact solutions have applied only to one-dimensional systems. But writing in *Physical Review Letters* Dukelsky et al.¹ introduce a new family of exactly solvable theoretical models dealing with quantum many-body systems in any number of dimensions: one, two, three or even more.

A wide range of many-body systems has been studied, with varying numbers of particles. For example, systems of nuclei ($n \approx 10^2$ nucleons) are important in nuclear physics, and atomic and molecular physics deal with systems involving $n \approx 10^2$ electrons, whereas mesoscopic ($n \approx 10^2$ – 10^6) and macroscopic ($n \approx 10^{23}$) systems arise in condensed-matter physics. Real systems are usually extraordinarily complex and involve a large set of parameters, most of which are irrelevant for discussing the phenomena under investigation. So theoretical physicists begin by defining a model — that is, an idealized system much simpler than the real one, but retaining all the necessary ingredients to discuss the physical properties of interest. However, there are only a few examples of 'exactly solvable' models of this type in which one can calculate exactly all the possible quantum states and their energies and/or the various thermodynamic quantities of the system. Such exact solu-

tions always play a decisive role in theoretical physics.

With very few exceptions, the exact solutions of many-body problems can be found only in one dimension. For example, the Tomonaga–Luttinger solution for a one-dimensional system of interacting fermions (particles such as electrons) demonstrates how the Landau–Fermi liquid description of electrons, which has been the theoretical basis of the whole quantum theory of solids, breaks down in one or two dimensions. This discovery had a considerable impact on condensed-matter physics, and gave rise to a new field concerned with the properties of the one-dimensional state known as the ‘Luttinger liquid’. In contrast, several exactly solvable models have been developed in nuclear physics that make use of group-theory techniques. These models are based on the concept of dynamical symmetries, and a non-trivial generalization of this approach is the ‘pairing model’, in which particles that form a pair interact through a force with a particular strength. Such a model has been applied to discussions of superfluidity, and also to studies of superconductivity in small metallic grains.

The pairing model was solved exactly by Richardson² more than 30 years ago, a result that passed almost unnoticed despite its enormous potential impact on nuclear as well as condensed-matter physics. In their paper, Dukelsky and colleagues¹ provide an important generalization of Richardson’s work, leading to new classes of pairing models that are exactly solvable in any dimension. To obtain this result, the authors apply group-theoretical and algebraic methods similar to those already used by Gaudin³. They discover three classes of integrable (a remarkable property that makes them mathematically tractable) pairing models, which they were able to solve exactly.

In analogy with the work of Gaudin, the three new classes of pairing models found by Dukelsky *et al.*¹ are named the rational, the trigonometric and the hyperbolic models. The authors go on to apply the rational model to a two-dimensional system of 18 fermions with repulsive interactions, for which they find an exact numerical solution. Their first results indicate that there are attractive correlations between pairs of fermions, despite the underlying repulsive interaction. Such models might apply to high-temperature superconductivity, in which mutually repulsive electrons have to pair up to flow without resistance. In further work, Dukelsky *et al.*⁴ propose an exact correspondence between the solution of the quantum pairing model and a two-dimensional classical electrostatic problem: the trigonometric model can be mapped to an electrostatic problem on a spherical surface, whereas the hyperbolic model

seems to correspond to a mapping on a more general surface.

The new solutions of the pairing models seem to be limited to systems of finite size — a necessary condition for solving equations giving exact quantum states. But they can certainly be applied to systems where n is quite large, from which the extrapolation to infinite n is fairly accurate. The solutions are also restricted to systems at a temperature of absolute zero (-273°C), but because the exact solution provides all the quantum states and their energies, any thermodynamic quantity can be calculated at a non-zero temperature by thermal averaging, according to the Boltzmann formula of statistical mechanics. Exact solutions, as well as predicting the particular properties of a simplified system, can also indicate how good a given approximation is to a many-body system.

In all these respects, the models of Dukelsky *et al.* are highly welcome. Another useful feature of these models is that they apply equally to boson and fermion systems. All known particles are either fermions (such as electrons) or bosons (such as photons), which obey different quantum statistics. As the authors show¹, the rational model may have some bearing on fermion systems, in particular the physics of high-temperature superconductors. In the case of boson systems, Dukelsky and Shuck⁵ have applied the rational model to the transition of a group of low-temperature atoms to a Bose–Einstein condensate — a state of matter in which all the atoms share the same quantum state. In similar work, Dukelsky and Pittel⁶ applied the model to a system of bosons with short-range repulsive interactions, and obtained new evidence for the validity of the ‘interacting boson model’ of nuclear physics.

The physics of strongly correlated fermion systems is one area of active research that should benefit immediately from these new families of exactly solvable models. Despite considerable theoretical efforts, mainly motivated by the problems raised by high-temperature superconductivity, such systems remain a huge challenge to theorists. These, and other seemingly intractable problems, would welcome any insight offered by a new family of theoretical tools. The results obtained by Dukelsky *et al.*¹ for a two-dimensional system of fermions with repulsive interactions is a first step in this direction. ■

Michel Héritier is in the Laboratoire de Physique des Solides d’Orsay, Centre Scientifique d’Orsay, 91405 Orsay, France.

e-mail: heritier@lps.u-psud.fr

1. Dukelsky, J., Esebbag, C. & Schuck, P. *Phys. Rev. Lett.* **87**, 066403 (2001).
2. Richardson, R. W. *Phys. Lett.* **3**, 277–279 (1963).
3. Gaudin, M. *J. Phys. (Paris)* **37**, 1087–1098 (1976).
4. Dukelsky, J. *et al.* *Phys. Rev. Lett.* (submitted).
5. Dukelsky, J. & Schuck, P. *Phys. Rev. Lett.* **86**, 4207–4210 (2001).
6. Dukelsky, J. & Pittel, S. *Phys. Rev. Lett.* **86**, 4791–4794 (2001).

Daedalus

Recycled oil

The internal-combustion engine, says Daedalus, is inherently dirty. A gas explosion or ignition occurs in a cold cylinder, and the molecular combustion chains are quenched on its cold walls. DREADCO engineers are now seeking to make the two-stroke diesel, the dirtiest of them all, into a clean motor.

Daedalus recalls one test diesel engine that could not be turned off because it was burning its own lubricating oil. Old technology used mixed oil and petrol to drive and lubricate a two-stroke engine. So Daedalus’s new engine will use the same oil first as lubricant and then as fuel.

A diesel engine is ideal for such experiments because it can burn almost anything. Whereas petrol has to be carefully formulated, the earliest diesel burnt coal dust, and almost any oil will power a diesel somehow. The ideal engine would have a cylinder so thick with oil that all combustion chains would be terminated by a liquid film. The fast-flowing lubricant would be rapidly dirtied, because all the pollution would end up in it. But instead of cleaning the oil, Daedalus wants to burn it.

The key component here is the cooling system. Every diesel engine has to have one — apart from the adiabatic diesel, that is, which just runs hot — and Daedalus proposes one large enough to cool not merely the engine block, but the exhaust gas as well. The engine will then discharge cooled filthy exhaust. Daedalus wants to flow this outgoing exhaust against incoming spent lubricating oil, possibly on a rotating disc or corrugated cylinder. If this is designed properly, the outgoing gas should yield up its nasties to the incoming oil. Carbon, other solid particulates, nitrogen oxides and similar pollutants would be taken up by the spent oil, leaving only clean air (still carrying water vapour and carbon dioxide) to leave the system. The dirty oil will feed the diesel. It will first lubricate, and then drive, the engine.

The new DREADCO engine will have a modest capacity, much less than a litre. Its cooling and auxiliary gear will be heavy, making it more suitable for urban vans than for motor cars. Its cleanliness and use of fuel as lubricant will be a bonus. But Daedalus hopes that the use of incoming fuel as a pollution-scavenger will be more widely adopted by internal-combustion engineers. It should give their engine another boost in its long fight against the encroaching fuel cell.

David Jones

Water capture by a desert beetle

This insect has a tailor-made covering for collecting water from early-morning fog.

Some beetles in the Namib Desert collect drinking water from fog-laden wind on their backs¹. We show here that these large droplets form by virtue of the insect's bumpy surface, which consists of alternating hydrophobic, wax-coated and hydrophilic, non-waxy regions. The design of this fog-collecting structure can be reproduced cheaply on a commercial scale and may find application in water-trapping tent and building coverings, for example, or in water condensers and engines.

The Namib Desert in southern Africa supports a unique sand-dune fauna² and experiences high winds, extreme daytime temperatures and dense, early-morning fog^{3,4}, but rainfall is so low and variable as to be considered negligible here⁵. The tenebrionid beetle *Stenocara* sp. (Fig. 1a) tilts its body forwards into the wind to collect water in a manner that is typical of the family³. Droplets form on the top (front) fused 'wings' (elytra) and roll down the beetle's surface to its mouthparts. The mechanism by which water is extracted from the air and formed into large droplets has so far not been explained, despite its biomimetic potential.

A similar rolling of water droplets occurs on lotus leaves, which repel raindrops on their uniformly superhydrophobic surface⁶. But a surface of this nature would not help the beetle — the fog water droplets encountered by *Stenocara* are much finer (1–40 µm diameter⁷) than raindrops and, without a special controlling mechanism, would quickly be lost to the heat and winds of the desert (water droplets in fog are so light that they travel almost horizontally in a breeze)^{7–9}. Water collection is therefore not as simple as it may seem.

On a macroscopic scale, the elytra of *Stenocara* are covered in a near-random array of bumps 0.5–1.5 mm apart, each about 0.5 mm in diameter (Fig. 1a). At the microscopic level, the peaks of these bumps are smooth, with no covering (Fig. 1b), whereas the troughs, including their sloping sides, are covered by a microstructure coated in wax (Fig. 1b). The microstructure consists of flattened hemispheres, 10 µm in diameter and arranged in a regular hexagonal array (Fig. 1c), creating a superhydrophobic system reminiscent of the lotus leaf.

The droplet-producing system works by 'growing' droplets on the hydrophilic seeding points of the peaks (as seen in video recordings). The water in the fog settles on the peaks, forming fast-growing droplets that bind to the elytra; water striking the hydrophobic slopes can also be collected as

it may bounce or be blown to a hydrophilic region. Each attached droplet eventually reaches a size at which its contact area covers the entire hydrophilic island. Beyond this size, the ratio of its mass to its surface-contact area increases rapidly until the capillary force that attaches it to the surface is overcome (this force is dictated by the area of the hydrophilic island). At this point, the droplet detaches and rolls down the tilted beetle's surface, guided by the slight purchase afforded by other peaks along its path.

By this stage, the droplet is sufficiently massive to roll into the wind, its cross-section (the area of contact with the wind) having increased far more slowly than its volume. Using the control volume form of Euler's first law¹⁰, it is possible to estimate the windspeed, v , below which the droplets will roll downwards:

$$v \approx \sqrt{\left[\left(\frac{\rho_{\text{water}}}{\rho_{\text{air}}} \right) \left(\frac{4}{3} \right) R g \sin(\theta) \right]}$$

where R is the droplet radius, g is the gravitational constant (9.8 m s^{-2}), θ is the tilt angle and ρ is the density of a medium ($\rho_{\text{air}} = 1 \text{ kg m}^{-3}$, $\rho_{\text{water}} = 1,000 \text{ kg m}^{-3}$). This predicts that the diameter of a spherical droplet must exceed 5 mm if it is to roll down a 45° incline into a headwind of 5 m s^{-1} (as measured in Namibia). This is fairly consistent with video footage of the beetles in their natural environment, in which droplets of roughly 4–5-mm diameter form in a steady, self-replenishing stream.

To determine the biomimetic potential of the structure, we partially embedded 0.6-mm glass spheres into microscope slides coated with warm wax. In one model, spheres were arranged in a square array with a spacing of 0.6 mm; in another, spheres were randomly arranged with an average spacing of 0.5 mm. For comparison, we used a uniformly hydrophobic surface of smooth wax and a uniformly hydrophilic surface of bare glass, which was cleaned by ultrasonic agitation in isopropyl alcohol for 15 min. The four samples were inclined at 45° and sprayed equally with a fine mist of water, which was collected at the base of each slide. We repeated this experiment 10 times using samples at 22°C .

The square array of beads collected the most water (1 unit), with large, almost spherical droplets quickly forming to roll downwards when they reached a diameter of 3.8–4.0 mm. The random array collected an average of 0.95 units and the bare wax sample an average of 0.5 units, as droplets of

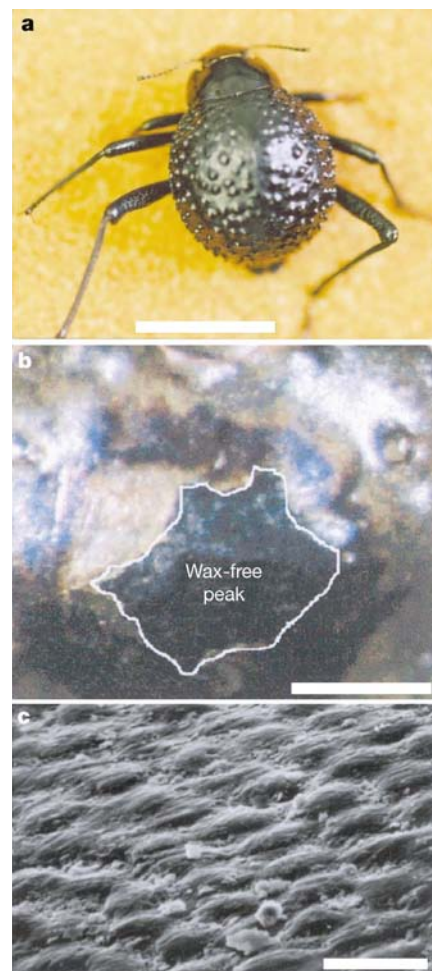


Figure 1 The water-capturing surface of the fused overwings (elytra) of the desert beetle *Stenocara* sp. **a**, Adult female, dorsal view; peaks and troughs are evident on the surface of the elytra. **b**, A 'bump' on the elytra, stained with Red O for 15 min and then with 60% isopropanol for 10 min, a procedure that tests for waxes. Depressed areas of the otherwise black elytra are stained positively (waxy, coloured), whereas the peaks of the bumps remain unstained (wax-free; black). **c**, Scanning electron micrograph of the textured surface of the depressed areas. Scale bars, **a**, 10 mm; **b**, 0.2 mm; **c**, 10 µm.

about 1 mm diameter; many droplets were lost when blown off the edges of the slide. Bare glass gave the most variable results: a single large (over 7 mm) flat droplet formed on the surface and ran down the slide by a specific route that varied within each experiment. If the route led to the collecting vial, as in four of the experiments, 1 unit of water was collected. If it did not, as in six experiments, no water was collected.

The results were virtually identical at substrate temperatures of 66°C to those at 22°C , except in the case of bare glass, where the flat droplets evaporated because of their large surface area. We conclude that the

combination of hydrophilic and hydrophobic points was best for collecting water from mist at both 22 °C and 66 °C.

The *Stenocara* structure¹¹ can easily be reproduced in sheet form by injection-moulding or printing techniques, allowing a variety of uncooled devices to be produced for controlled collection of vapour, including water for drinking or farming in inhospitable regions^{7,9} (see supplementary information).

Andrew R. Parker*, Chris R. Lawrence†

*Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

e-mail: andrew.parker@zoo.ox.ac.uk

†Mechanical Sciences Sector, QinetiQ,

Farnborough GU14 0LX, UK

1. Hamilton, W. J. & Seely, M. K. *Nature* **262**, 284–285 (1976).
2. Koch, C. *Ann. Trans. Mus.* **24**, 61–106 (1962).
3. Crawford, C. S. *Biology of Desert Invertebrates* (Springer, Berlin, 1981).
4. Roberts, C. S., Seely, M. K., Ward, D., Mitchell, D. & Campbell, J. D. *Physiol. Entomol.* **16**, 463–475 (1991).
5. Louw, G. N. *Symp. Zool. Soc. Lond.* **31**, 297–314 (1972).
6. Watanabe, K. J. *Jap. Soc. Tribol.* **45**, 354–359 (2000).
7. <http://www.unep.or.jp/ietc/publications/techpublications/techpub-8a/fog.asp>.
8. Machin, J. J. *Exp. Biol.* **53**, 753–762 (1970).
9. <http://www.oas.org/usde/publications/unit/oea59e/ch12.htm>
10. McGill, D. J. & King W. W. *Statics and an Introduction to Dynamics* 2nd edn (PWS-Kent, Boston, 1989).
11. UK Patent Application no. 0109814.4, filed 23/04/01.

Supplementary information accompanies the paper on *Nature's* website (www.nature.com).

Metabolism

Partial leptin deficiency and human adiposity

The adipocyte-derived hormone leptin is crucial for energy homeostasis in mammals; mice¹ and humans^{2,3} without it suffer from a voracious appetite and extreme obesity. The effect on energy balance of variations in plasma leptin above a minimal threshold is uncertain, however, particularly in humans. Here we examine a group of individuals who are genetically partially deficient in leptin, and show that differences in circulating leptin levels within the range found in normal human populations can directly influence the laying down of fat tissue (adiposity).

In three unrelated families of Pakistani origin, we identified 13 subjects who were heterozygous for a frameshift mutation (deletion of a glycine at residue 133; Δ G133) in the *ob* gene, which encodes leptin. We also identified three of their relatives as homozygous for the wild-type *ob* sequence. Serum leptin levels and anthropometric measurements in these subjects were compared with those in 96 ethnically matched controls with a similar sex distribution (38% and 44% male, respectively) and age (41 ± 12 yr and 48 ± 15 yr, respectively).

Serum leptin concentrations in Δ G133 heterozygotes were markedly lower than in controls (Fig. 1a). As expected, these were positively correlated with body-mass index (BMI) in control subjects ($r = 0.54$) and in the wild-type relatives of the heterozygotes (Fig. 1b). In contrast, there was no significant correlation between BMI and serum leptin in the Δ G133 heterozygotes.

In all human populations studied so far, serum leptin is positively correlated with indices of adiposity⁴. However, lower leptin levels in Δ G133 heterozygotes were accompanied by an increased prevalence of obesity, with 76% of heterozygotes having a BMI greater than 30 kg m^{-2} , compared with 26% of controls ($P < 0.01$). As BMI is

only a surrogate measure of body fat, we measured body fat directly by dual-energy X-ray absorptiometry in 12 of the Δ G133 heterozygotes and in 6 Pakistani subjects who were wild-type at the leptin locus. We compared these direct measures of body fat to body-fat predictions on the basis of age, sex, height and weight⁵.

In wild-type subjects, the mean measured percentage of body fat was similar to the predicted value (Fig. 1c). In Δ G133 heterozygotes, however, it significantly exceeded the predicted proportion of body fat (41.4% and 34.4%, respectively; $P < 0.001$), with all 12 heterozygous individuals showing a deviation in the same direction.

Humans who inherit only one functional copy of the leptin gene therefore have markedly lower serum leptin levels than control subjects, particularly in view of their

increased mass of adipose tissue. These observations are supported by studies of leptin deficiency in rodents — heterozygous *Lep^{ob/+}* mice have about 32% more fat and adjusted serum leptin concentrations that are about 27% lower than their wild-type littermates⁶, which is strikingly similar to our findings in humans.

If leptin evolved principally to signal the transition between states of starvation and nutritional adequacy⁷, then sensitivity to changes in its circulating concentration should be enhanced at very low values. This could account for inconsistent effects on appetite or weight when leptin in normal or obese humans is supplemented by injection of recombinant human leptin⁸. If the human energy-balance system is insensitive to positive perturbations of leptin, is it also insensitive to negative ones? Our results would suggest that this is not the case.

Our findings show that a relatively small drop in leptin production may be sensed by the homeostatic feedback system that controls energy balance, with fat mass being increased in an attempt to restore leptin levels to some 'set point'. As a substantial minority of subjects with common forms of obesity, not associated with leptin mutations, have relatively low levels of circulating leptin⁹, our results indicate that augmenting serum leptin in this subgroup may be therapeutically worthwhile.

I. Sadaf Farooqi*, Julia M. Keogh*, Sri Kamath†, Sarah Jones‡, William T. Gibson§, Rebecca Trussell||, Susan A. Jebb‡, Gregory Y. H. Lip†, Stephen O'Rahilly*

*University Departments of Medicine and Clinical

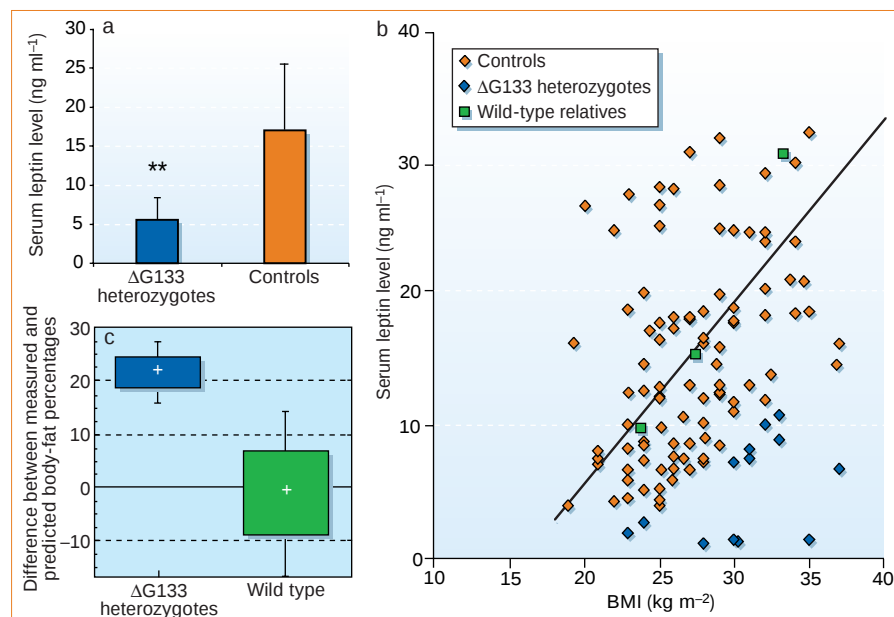


Figure 1 Human heterozygotes for the Δ G133 mutation in the leptin gene have low serum leptin levels and increased body-fat percentage relative to controls. **a**, Serum leptin concentrations (mean $\pm$ s.d.) in 13 Pakistani subjects who are heterozygous for the Δ G133 leptin mutation and in 96 control subjects from the same ethnic group; ** denotes $P < 0.001$. **b**, Serum leptin concentrations plotted against body-mass index (BMI), showing the regression line for controls; correlation coefficient, 0.54. **c**, Body-fat percentage, measured by dual-energy X-ray absorptiometry in 12 Δ G133 heterozygotes and 6 wild-type subjects, compared to that predicted by the Deurenberg formula⁵ for both groups. Values are means $\pm$ 2 s.e.m.

Biochemistry, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK
e-mail: sorahill@hgmp.mrc.ac.uk
†University Department of Medicine, City Hospital, Dudley Road, Birmingham B18 7QH, UK
‡MRC Human Nutrition Research, Fulbourn Road, Cambridge CB1 9NL, UK
Departments of §Medical Genetics and ||Paediatric Endocrinology, Alberta Children's Hospital, 18200 Richmond Road Southwest, Alberta, Calgary T2T 5C7, Canada

1. Zhang, Y. et al. *Nature* **372**, 425–432 (1994).
2. Montague, C. T. et al. *Nature* **387**, 903–908 (1997).
3. Farooqi, I. S. et al. *N. Engl. J. Med.* **341**, 879–884 (1999).
4. Havel, P. J. *Proc. Nutr. Soc.* **59**, 359–371 (2000).
5. Deurenberg, P., Weststrate, J. A. & Seidell, J. C. *Br. J. Nutr.* **65**, 105–114 (1991).
6. Chung, W. K. et al. *Am. J. Physiol.* **274**, 985–990 (1998).
7. Flier, J. S. *J. Clin. Endocrinol. Metab.* **83**, 1407–1413 (1998).
8. Heymsfield, S. et al. *J. Am. Med. Assoc.* **282**, 1568–1575 (1999).
9. Ravussin, E. et al. *Nature Med.* **3**, 238–240 (1997).

Bird migration

Magnetic cues trigger extensive refuelling

Long stretches of sea and desert often interrupt the migration routes of small songbirds, whose fat reserves must be restored before these can be crossed as they provide no opportunity for refuelling. To investigate whether magnetic cues might enable inexperienced migratory birds to recognize a region where they need to replenish their body fat, we caught and held thrush nightingales (*Luscinia luscinia*) in Sweden just before their first migration and exposed them to a magnetic field simulating that at a migratory stopover in northern Egypt, before the Sahara Desert. We found that this magnetic field stimulated the birds to extend their fat-deposition period, indicating that magnetic cues may help small migratory birds to confront large ecological barriers.

Long-distance migration is common in birds, with some species migrating for more than 6 months a year¹. Many rely on circannual rhythms, which are fine-tuned by photoperiod, for the timing of activities such as breeding, migration and moulting². Most songbirds migrate alone in a series of nocturnal flights, using both celestial cues and information from the Earth's magnetic field to select and maintain migratory direction³. However, evidence of true magnetic navigation in birds remains circumstantial⁴. The loggerhead sea turtle (*Caretta caretta*) is sensitive to both the inclination angle and the intensity of the Earth's magnetic field⁵ (sea-turtle hatchlings respond to different magnetic fields by swimming in directions that keep them within a specific region of the north Atlantic Ocean).

The main fuel used by migrating birds is fat, which is laid down at stopover sites en route¹. As carrying large fat stores incurs

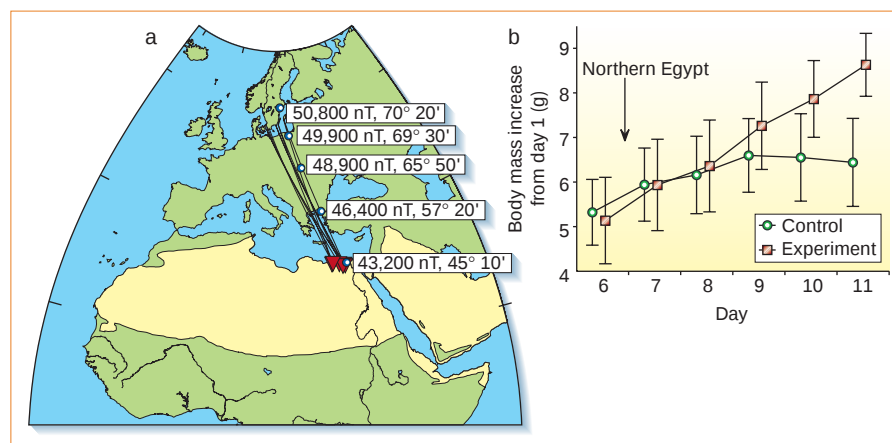


Figure 1 Effect of magnetic-field simulation on migratory refuelling in thrush nightingales (*Luscinia luscinia*). **a**, Map showing the points of autumn recovery of migrating birds originally ringed in Sweden ($n=9$) and the stopover sites for which magnetic fields were simulated (total intensity (values in nanotesla) and inclination). **b**, Average body-mass increase ($\pm$ s.e.) during the experiment (registered automatically using a Precisa Balance 310C). The field experienced by experimental birds was changed to that of four localities during the course of the experiment, calculated according to IGRF2000 (ref. 12). Experimental birds remained from day 7 in the magnetic field of northern Egypt until the end of the experiment at day 11, when all birds were released back into the wild. The magnetic system consisted of two independent series of four quadratric coils each, arranged orthogonally¹³. We verified the homogeneity of the magnetic field and set the vertical and horizontal components for the stopover sites using a Zeiss Jena theodolite with a fluxgate magnetometer (Bartington Instruments) and a proton magnetometer (GEM Systems). Further details are available from the authors.

increases in flight cost and predation risk^{6,7}, most bird species accumulate only small fat deposits (20–30% of lean body mass) and refuel at several successive stopover sites⁶. But to cross large ecological barriers such as the Gulf of Mexico or the Sahara Desert, fuel loading needs to be much greater — some birds have been found to double their mass⁸ before crossing the Sahara, which involves flight distances of at least 1,500 km.

The extent and timing of refuelling has been assumed to be governed by the circannual rhythm⁹. But as there is a variation in the timing of migration onset in most populations (for example, because of variation in time of breeding), as well as in weather and feeding conditions during migration, a bird cannot gauge its latitudinal position from purely seasonal parameters. Other external cues may be necessary to indicate where large fuel loads need to be accumulated.

Autumn recoveries of thrush nightingales ringed in Sweden at the start of their migration showed that they congregate in northern Egypt¹⁰ (Fig. 1a), presumably in preparation for crossing the Sahara Desert. To investigate whether the rate of fat deposition in thrush nightingales during this migratory stopover could be influenced by the magnetic field in northern Egypt, we randomly assigned first-year birds, caught at Tovetorp Zoological Research Station, Sweden, in August 2000, to either a magnetic treatment, in which the magnetic field was gradually changed to that of northern Egypt, or a control treatment, in which birds experienced the ambient magnetic field.

Birds subjected to the changing magnetic field had a larger increase in body mass than control birds as a result of extending their fuelling period once the magnetic field

became the same as that in Egypt (Fig. 1b). Experimental birds increased in mass by 3.5 ± 0.8 s.e. ($n=8$) from days 6 to 11, whereas control birds increased in mass by 1.1 ± 0.6 s.e. ($n=8$) ($F_{4,48}=4.40$, $P=0.004$). No differences were observed between replications ($F_{1,12}=0.93$, $P=0.35$).

We have discovered a surprising external cue that helps to optimize this bird's chances of successful migration, and which works in concordance with orientation behaviour and endogenous rhythm¹¹ to provide precise information about geographical position when such information is crucial. We cannot yet say whether the fat-deposition response described here is an evolved response to the magnetic field in a specific area, or whether the birds are reacting to the latitudinal change in magnetic field.

Thord Fransson*†, **Sven Jakobsson***, **Patrik Johansson‡**, **Cecilia Kullberg***, **Johan Lind***, **Adrian Vallin***

*Department of Zoology, Stockholm University, SE-106 91 Stockholm, Sweden

†Swedish Museum of Natural History, Bird Ringing Centre, SE-104 05 Stockholm, Sweden
e-mail: thord.fransson@nrm.se

‡Geological Survey of Sweden, SE-751 28 Uppsala, Sweden

1. Alerstam, T. *Bird Migration* (Cambridge Univ. Press, Cambridge, 1990).
2. Gwinner, E. *Ibis* **138**, 47–63 (1996).
3. Able, K. P. *Trends Ecol. Evol.* **8**, 367–371 (1993).
4. Berthold, P. *Control of Bird Migration* (Chapman & Hall, London, 1996).
5. Lohmann, K. J. & Lohmann, C. M. F. *Nature* **380**, 59–61 (1996).
6. Alerstam, T. & Lindström, Å. in *Bird Migration: The Physiology and Ecology* (ed. Gwinner, E.) 331–351 (Springer, Berlin, 1990).
7. Kullberg, C., Fransson, T. & Jakobsson, S. *Proc. R. Soc. Lond. B* **263**, 619–624 (1996).
8. Fry, C. H., Ash, J. S. & Ferguson-Lees, I. J. *Ibis* **112**, 58–82 (1970).
9. Berthold, P. *Trends Ecol. Evol.* **6**, 254–257 (1991).

10. Swedish Museum of Natural History, Bird Ringing Centre. *Annual Reports of Swedish Bird Ringing 1960–98* (ISSN 0282-390X).
11. Thorup, K. & Rabel, J. *J. Avian Biol.* **32**, 111–119 (2001).
12. International Association of Geomagnetism and Aeronomy. *Geophys. J. Int.* **141**, 259–262 (2000).
13. Lohmann, K. & Lohmann, C. M. F. *J. Exp. Biol.* **194**, 23–32 (1994).

COMMUNICATIONS ARISING

Tumour suppressors

Effect of DNA damage on a BRCA1 complex

The tumour-suppressor protein BRCA1 mediates its biological functions by interacting with cellular factors^{1,2} such as the CtIP polypeptide^{3,4}, a substrate for the ATM (for 'ataxia telangiectasia mutated') protein kinase⁵. Li *et al.*⁶ report that the BRCA1–CtIP interaction is disrupted by ionizing radiation and by other genotoxic stresses that induce phosphorylation of CtIP by ATM kinase, and that this dissociation of the BRCA1–CtIP complex in turn modulates the transcription of DNA-damage-response genes⁶. We have shown that the BRCA1-binding domain of CtIP (amino-acid residues 133–369) is distal to the sites that are phosphorylated by ATM kinase (residues S664 and S745)⁷. We now show that the BRCA1–CtIP complex is stable in irradiated cells, and that the phosphorylated isoforms of CtIP that are induced by ionizing radiation still interact *in vivo* with BRCA1. We conclude that disruption of the BRCA1–CtIP complex cannot account for induction of DNA-damage-response genes in the way proposed by Li *et al.*⁶.

To investigate the effect of genotoxic stress on CtIP, we treated human T24 carcinoma cells with ionizing radiation or ultraviolet light, and immunoblotted the cell lysates with a CtIP-specific monoclonal antibody⁷. As expected, ionizing radiation induced the formation of the phosphorylated CtIP isoforms⁶ (Fig. 1a, top, lane 3). These species migrate more slowly than CtIP polypeptides from untreated cells (lane 1) and are converted to a faster-migrating form after incubation with λ -phosphatase (lane 4). As expected, BRCA1 was also hyperphosphorylated in cells exposed to ionizing radiation (Fig. 1a, bottom)^{8,9}.

To determine the effect of ionizing radiation on the BRCA1–CtIP complex, we immunoprecipitated T24-cell lysates with a BRCA1-specific polyclonal antiserum, and monitored each precipitate for CtIP by immunoblotting with a CtIP-specific monoclonal antibody. As expected, CtIP was detected in BRCA1 immunoprecipitates from lysates of untreated and ultraviolet-irradiated cells (Fig. 1b, top, lanes 5, 9), which is consistent with our previous findings⁷ but not with those of Li *et al.*¹⁰.

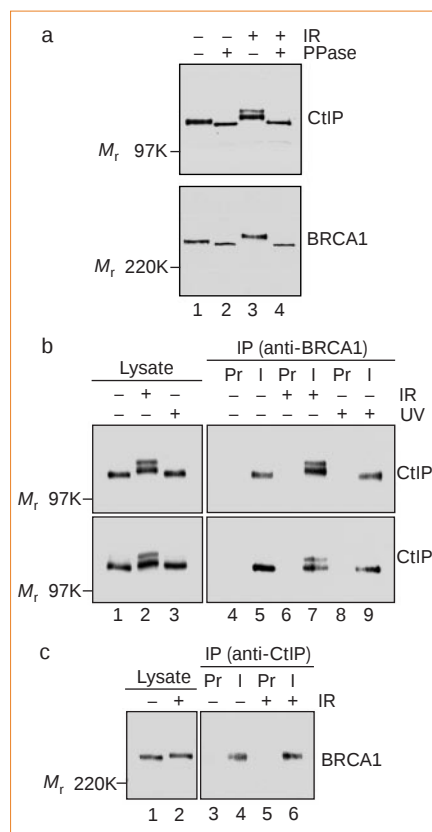


Figure 1 Association of BRCA1 and CtIP in irradiated cells. **a**, Induction of CtIP phosphorylation. Lysates from T24 carcinoma cells (lanes 1, 2, untreated; lanes 3, 4, treated with 40 Gy ionizing radiation (IR) were immunoblotted with CtIP-specific (top) or BRCA1-specific (bottom) monoclonal antibodies⁷; the indicated lysates were pretreated with λ -phosphatase (PPase). **b**, Co-immunoprecipitation of CtIP with BRCA1. T24 cells (top) and GM00637H fibroblasts (bottom) were irradiated with 40 Gy IR or 10 J m⁻² ultraviolet (UV) light, and lysates were immunoblotted for CtIP (lanes 1–3). Alternatively, lysates were immunoprecipitated with BRCA1-specific antiserum⁷ (I) or pre-immune serum (Pr) and then immunoblotted for CtIP (lanes 4–9). **c**, Co-immunoprecipitation of BRCA1 with CtIP. Lysates of HBL100 epithelial cells were either immunoblotted for BRCA1 (lanes 1, 2) or immunoprecipitated with CtIP-specific 210 antiserum⁷ and then immunoblotted for BRCA1 (lanes 3–6).

Moreover, we recovered all CtIP species, including the hyperphosphorylated forms, in BRCA1 immunoprecipitates prepared from cells exposed to ionizing radiation (Fig. 1b, lane 7). The same results were obtained in all cell lines tested, including the two lines examined by Li *et al.*⁶: T24 carcinoma cells (Fig. 1b, top) and SV40-transformed GM00637H fibroblasts (Fig. 1b, bottom).

The stability of the CtIP–BRCA1 complex was also evident in reciprocal co-immunoprecipitation experiments, in which comparable amounts of BRCA1 polypeptides were present in CtIP immunoprecipitates from both untreated cells and cells exposed to ionizing radiation (Fig. 1c, lanes 4, 6). Our results indicate that, contrary to the findings of Li *et al.*, the CtIP–BRCA1 complex is stable to genotoxic stress such as ultraviolet or ionizing radiation.

Foon Wu-Baer, Richard Baer

Institute of Cancer Genetics and Department of Pathology, Columbia University College of Physicians and Surgeons, 1150 St Nicholas Avenue, New York, New York 10032, USA
e-mail: rb670@columbia.edu

1. Scully, R. & Livingston, D. M. *Nature* **408**, 429–432 (2000).
2. Deng, C.-X. & Brodie, S. G. *BioEssays* **22**, 728–737 (2000).
3. Yu, X., Wu, L. C., Bowcock, A. M., Aronheim, A. & Baer, R. *J. Biol. Chem.* **273**, 25388–25392 (1998).
4. Wong, A. K. *et al.* *Oncogene* **17**, 2279–2285 (1998).
5. Kim, S. T., Lim, D. S., Canman, C. E. & Kastan, M. B. *J. Biol. Chem.* **274**, 37538–37543 (1999).
6. Li, S. *et al.* *Nature* **406**, 210–215 (2000).
7. Yu, X. & Baer, R. *J. Biol. Chem.* **275**, 18541–18549 (2000).
8. Scully, R. *et al.* *Cell* **90**, 425–435 (1997).
9. Thomas, J. E., Smith, M., Tonkinson, J. L., Rubinfeld, B. & Polakis, P. *Cell Growth Differ.* **8**, 801–809 (1997).
10. Li, S. *et al.* *J. Biol. Chem.* **274**, 11334–11338 (1999).

Li *et al.* reply — Wu-Baer and Baer confirm our original observation that CtIP is phosphorylated in an ATM-dependent manner in response to γ -radiation. We have shown that phosphorylation by ATM kinase of CtIP at serine residues 664 and 745 is required to liberate DNA-damage-response genes such as *GADD45* from repression. This is consistent with our more recent finding that overexpression in mammalian cells of a phosphorylated CtIP mutant with a double alanine substitution at serines 664 and 745 disrupts the radiation-induced cell-cycle checkpoint between G₂ and M phases. The functional consequence of radiation-induced, ATM-dependent phosphorylation of CtIP is therefore clear.

With respect to the mechanism that underlies this process, we proposed that phosphorylation of CtIP leads to its dissociation from BRCA1, freeing BRCA1 to participate in the activation of DNA-damage-response genes. We do, however, appreciate the potential for experimental artefact in using only soluble co-immunoprecipitation to detect protein–protein interactions, particularly when different antibodies and cell lines are used. (We used human colon cancer cell line HCT 116 and human fibroblasts GM09607A and GM00637G, whereas Wu-Baer and Baer used human bladder carcinoma cell line T24 and fibroblast GM00637H.)

Wu-Baer and Baer report reciprocal co-immunoprecipitation of BRCA1 and CtIP using two different cell lines instead of the same cells. Their observations are inconsistent with ours where the interaction status of BRCA1 and CtIP after γ -irradiation is concerned. This discrepancy should eventually be resolved by systematic investigation using an alternative and complementary method.

Shang Li, Nicholas S.Y. Ting, Lei Zheng, Phang-Lang Chen, Wen-Hwa Lee

Department of Molecular Medicine, University of Texas Health Science Center at San Antonio, Institute of Biotechnology, San Antonio, Texas 78245-3207, USA
e-mail: leew@uthscsa.edu

Energetic optimization of ion conduction rate by the K⁺ selectivity filter

João H. Morais-Cabral*, Yufeng Zhou & Roderick MacKinnon

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The K⁺ selectivity filter catalyses the dehydration, transfer and rehydration of a K⁺ ion in about ten nanoseconds. This physical process is central to the production of electrical signals in biology. Here we show how nearly diffusion-limited rates are achieved, by analysing ion conduction and the corresponding crystallographic ion distribution in the selectivity filter of the KcsA K⁺ channel. Measurements with K⁺ and its slightly larger analogue, Rb⁺, lead us to conclude that the selectivity filter usually contains two K⁺ ions separated by one water molecule. The two ions move in a concerted fashion between two configurations, K⁺-water-K⁺-water (1,3 configuration) and water-K⁺-water-K⁺ (2,4 configuration), until a third ion enters, displacing the ion on the opposite side of the queue. For K⁺, the energy difference between the 1,3 and 2,4 configurations is close to zero, the condition of maximum conduction rate. The energetic balance between these configurations is a clear example of evolutionary optimization of protein function.

The KcsA K⁺ channel contains multiple K⁺ ions in its conduction pore (Fig. 1a). One ion resides at the centre of a water-filled cavity located halfway across the membrane, and others are present in a narrow, 12 Å-long segment adjacent to the extracellular pore opening¹. This narrow segment is the selectivity filter where K⁺ ions, but not Na⁺ ions, bind in a nearly dehydrated state.

The selectivity filter is the central structural element that defines a K⁺ channel; its amino-acid sequence is conserved in all K⁺ channels throughout nature (Fig. 1b). In the filter, four strands of sequence TVGYG, one contributed by each of the four subunits, are arranged with their carbonyl oxygen atoms pointed inward towards the ion conduction pathway. This arrangement creates four potential ion-binding sites into which a K⁺ ion can bind in an essentially dehydrated state, surrounded by eight oxygen atoms from the protein. Here, we refer to these selectivity filter sites as positions 1–4, from outside to inside (Fig. 1b). The three outermost sites (positions 1–3) are formed exclusively by carbonyl oxygen atoms, whereas position 4, adjacent to the central cavity, is formed by four carbonyl oxygen atoms and four threonine side-chain oxygen atoms. In each of the four sites, a K⁺ ion can potentially be held at the centre of a box, with one plane of four oxygen atoms above and one plane below the ion.

Potassium ions diffuse through the channel at rates approaching 10⁸ ions s^{−1} under physiological conditions. To catalyse such high diffusion rates, the selectivity filter must allow a K⁺ ion to dehydrate, enter and cross the selectivity filter within about 10 ns. The aim of this study is to deduce the individual steps in translocation of the K⁺ ion across the selectivity filter, to understand how this process can occur at such a high rate. We carried out electrophysiological and X-ray crystallographic measurements in parallel to determine ion-throughput rates and ion distributions in the selectivity filter over a wide range of ion concentrations. Our ability to reach a plausible mechanism for conduction through the filter comes from a comparative analysis of K⁺ and Rb⁺. Rb⁺ is known to permeate K⁺ channels, but its conduction features are unusual^{2,3}. We find that the unusual properties of Rb⁺ conduction are mirrored by an anomalous distribution of Rb⁺ compared with K⁺ in the selectivity filter. This anomalous distribution provides the key to understanding the detailed mechanism of K⁺ conduction.

Rubidium ion

Single KcsA K⁺ channels were reconstituted into planar lipid bilayers, and ionic currents were recorded in the presence of equal concentrations of RbCl on both sides of the membrane. Figure 2a shows the current measured at 180 mV as a function of Rb⁺ concentration. The current increases steeply at low Rb⁺ concentrations (steep regime), begins to level off between 30 and 50 mM, and then continues to increase in a nearly linear fashion (linear regime).

X-ray crystallography was used to estimate the distribution of Rb⁺ ions in the selectivity filter and to investigate how the distribution changes as a function of Rb⁺ concentration. Crystals of the KcsA K⁺ channel (space group C2) were grown in solutions of varying Rb⁺ concentration. At concentrations less than 150 mM, we replaced Rb⁺ with Na⁺ or *N*-methylglucamine (NMG) to maintain constant ionic strength. Data were collected on flash-frozen crystals at the synchrotron, and difference Fourier omit maps were calculated from data to 3.4 Å Bragg spacings (see Methods). One-dimensional maps of electron density sampled along the axis of the selectivity filter (electron density profiles) show clear maxima

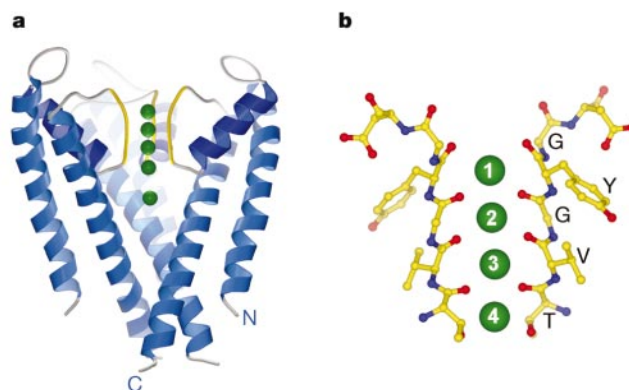


Figure 1 Binding sites for K⁺ ions in the KcsA K⁺ channel. **a**, Ribbon representation of the KcsA channel with the subunit closest to the viewer removed. Potassium ions (green spheres) bind at four locations in the selectivity filter (yellow) and in the water-filled cavity at the membrane centre (bottom ion). **b**, Close-up view of the selectivity filter in ball-and-stick representation, with the front and back subunits removed. The four K⁺ ions are numbered to indicate the location of binding sites in the filter; position 1 is closest to the extracellular solution and position 4 is closest to the cavity. Key amino acids forming the selectivity filter are shown. Pictures were prepared with GL-Render^{26,27}.

* Present address: Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520, USA.

and minima (Fig. 2b). At the lowest concentration of Rb^+ (5 mM), there is a strong peak at each end of the selectivity filter, corresponding to positions 1 and 4, with a small peak near position 3. As the Rb^+ concentration is raised, the profile changes: the peak near position 1 nearly doubles in size, a large peak appears near position 3, while the peak near position 4 actually diminishes in size (Fig. 2b and c). No electron density peak appears at position 2 in the filter. Once a Rb^+ concentration of about 30 mM is reached, there is little further change in the electron density profile except for a gradual increase in the peak at position 4.

The first interesting aspect of the electron density profiles is that their dependence on ion concentration correlates well with the concentrations over which the steep and linear regimes of the conduction–concentration curve are observed. That is, the transition from the low-concentration profile (a peak at positions 1 and 4) to the high-concentration profile (large peaks at positions 1 and 3 with a smaller peak at position 4) matches the transition from steep to linear conduction regimes (Fig. 2a and c). The second interesting aspect is that the electron density transition from low (5 mM) to high (≥ 30 mM) Rb^+ concentration clearly reflects an increase in the number of Rb^+ ions in the filter. This increase might at first seem to be a transition from two to three Rb^+ ions, but it is in fact more compatible with a transition from one to two ions. To see why this is so, imagine that when a single Rb^+ ion is in the filter it resides at either position 1 or 4, resulting in an occupancy of about 0.5 at each position, and that when two ions are in the filter they reside at positions 1 and 3, with an occupancy of 1.0 at each position. In this case, as the Rb^+ concentration is increased from low to high, and a growing fraction of channels go from having one ion to two ions in

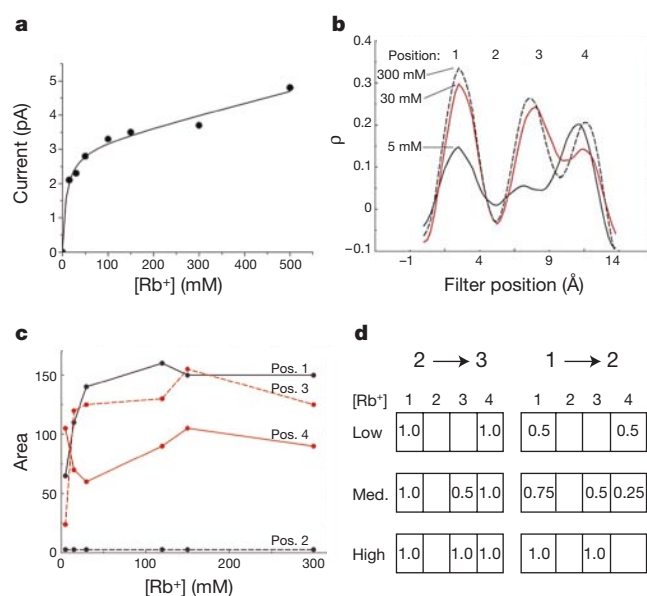


Figure 2 Rubidium conduction and crystallographic distribution of Rb^+ ions in the selectivity filter. **a**, Currents recorded from single KcsA K^+ channels in the presence of equal concentrations of internal and external RbCl are shown as a function of ion concentration. Recording voltage was +180 mV. **b**, Electron density (ρ) was sampled along the axis of the selectivity filter from difference Fourier omit maps (3.4 $\text{\AA}$ resolution, C2 crystal form) and is represented as one-dimensional profiles. Zero on the x-axis corresponds to the level of the α -carbon of residue 79 (extracellular entryway to the filter). Profiles are shown for crystals grown in the presence of 5, 30 and 300 mM RbCl . Filter positions 1–4 are indicated. **c**, The area of peaks in the electron density profiles corresponding to selectivity filter positions 1, 2, 3 and 4 are shown as a function of the Rb^+ concentration in which the crystals were grown. **d**, Sets of boxes representing the four positions in the selectivity filter, with the Rb^+ ion occupancy shown for low, intermediate (med.) and high Rb^+ concentrations. Expected occupancies for transitions from one to two ions (right) and from two to three ions (left) are shown (see text).

the filter, occupancy at position 1 will increase from 0.5 to 1.0, occupancy at position 3 will increase from 0 to 1.0, and occupancy at position 4 will decrease from 0.5 to 0 (Fig. 2d, right). These changes in occupancy are in qualitative agreement with the concentration dependence of peaks in the electron density profiles (Fig. 2b and c). The increase at position 1 and decrease at position 4 observed in the electron density is important because it is not explained by the transition from two to three ions. Such a transition would require positions 1 and 4 to each have an occupancy of 1.0 at low Rb^+ concentrations; as Rb^+ concentration is raised, an increasing fraction of channels with three ions in the filter would be manifest as an increasing occupancy at position 3 and nothing more (Fig. 2d, left). This simple explanation captures the approximate features of the data. More detailed considerations (below) account for the fact that the peak at position 4 does not disappear altogether, and actually becomes a little larger as Rb^+ concentration is increased beyond 30 mM (Fig. 2b and c).

Potassium ion

Potassium current through the KcsA K^+ channel is shown in Fig. 3a. As in the case of Rb^+ , K^+ conduction exhibits a steep regime at low ion concentrations and a linear regime at high concentrations. The striking difference between Rb^+ and K^+ is in the slope of the linear regime: the K^+ current continues to increase with ion concentration, whereas the Rb^+ current nearly levels off at a much smaller maximal value. This difference between Rb^+ and K^+ conduction has been documented previously and is a fairly general feature of K^+ channels^{2,3}.

To study K^+ ions in the selectivity filter, we grew crystals in the presence of various K^+ concentrations. Two crystal forms (space groups C2 and I4) were analysed, and they yielded similar results (see Methods). Electron density profiles (I4 crystal form, 2.4 $\text{\AA}$ Bragg spacings) for the concentration range 3–200 mM K^+ are shown in Fig. 3b. The 3 mM K^+ profile is similar to the 5 mM Rb^+

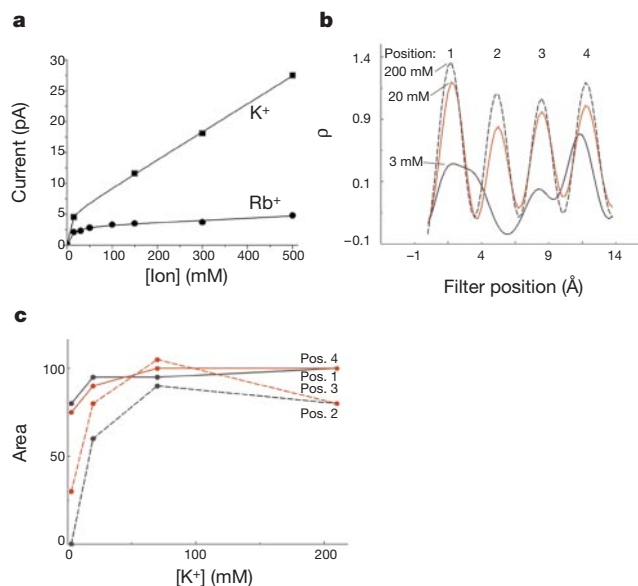


Figure 3 Potassium conduction and crystallographic distribution of K^+ ions in the selectivity filter. **a**, Currents recorded from single KcsA K^+ channels in the presence of equal concentrations of internal and external KCl and RbCl are shown as a function of ion concentration. Recording voltage was +180 mV. **b**, Electron density (ρ) was sampled as for Rb^+ -containing crystals, but the maps were calculated at 2.4 $\text{\AA}$ resolution from the I4 crystal form (see Methods). One-dimensional electron density profiles are shown for crystals grown in the presence of 3, 20 and 200 mM KCl . **c**, The area of peaks in the electron density profiles corresponding to selectivity filter positions 1, 2, 3 and 4 are shown as a function of the K^+ concentration in which the crystals were grown.

profile, showing a predominant peak on each end of the selectivity filter and a small peak near position 3 (Figs 2b and 3b). Therefore, Rb^+ and K^+ seem to be similar when only a single ion is present in the filter. However, as the K^+ concentration is raised, a very significant difference between the Rb^+ and K^+ profiles emerges. Instead of the appearance of a large peak near position 3, as in Rb^+ , two peaks appear at positions 2 and 3, yielding a profile with four roughly equal peaks. For both Rb^+ and K^+ , the transition from a low-concentration profile to a high-concentration profile occurs over a similar concentration range, 5–30 mM. Using the C2 crystal form, we found that the high-concentration profile in K^+ , showing four approximately equal peaks, was unchanged in K^+ concentrations of up to 500 mM (not shown).

Adjacent peaks in the K^+ electron density profile are separated by about 3.2 Å (Fig. 3b). Potassium ions have a diameter of 2.7 Å, so they could, in principle, fit in the filter side by side, but this would seem to be an unstable binding configuration for electrostatic reasons. A survey of a database of small-molecule structure (Cambridge Crystallographic Data Centre, <http://www.ccdc.cam.ac.uk>) showed us that two K^+ ions only very rarely occur with a separation distance of less than 3.5 Å. Therefore, although K^+ ions may in principle occupy adjacent sites in the selectivity filter, we expect that most often they are separated by an intervening water molecule.

When comparing conduction data from channels in a bilayer with structural data from channels in a crystal, an important assumption is made: that ion distributions measured in the absence of a driving force (at equilibrium in the crystal) are relevant to currents measured in the presence of a driving force (with a voltage across the bilayer). The weak voltage dependence of Rb^+ and K^+ conductance (current per voltage) justifies this assumption. That is, the conductance measured at 180 mV is, within a reasonably small factor, equal to the conductance extrapolated to 0 mV, where the steady-state ion distribution in the pore must approach the equilibrium distribution. Consequently, the general features of the conduction–concentration curve for K^+ channels, including KcsA, are similar near 0 and 200 mV for both K^+ and Rb^+ (ref. 3).

Conduction mechanism

The key experimental observations in this study include the following. First, the electron density profile along the selectivity filter for both Rb^+ and K^+ ions undergoes a transition as the ion concentration is raised. The transition clearly demonstrates an increase in the number of ions in the selectivity filter, and, under physiological ion concentrations (>100 mM K^+), the filter operates by a multi-ion conduction mechanism. The transition is also correlated with a change in the conduction–concentration curve from a steeply rising regime at low ion concentrations to a shallower, roughly linear regime at high ion concentrations. Second, at low ion concentrations Rb^+ and K^+ have similar electron density profiles, but at higher concentrations they are different. Rb^+ ions seem to prefer sites 1 and 3 in the filter, whereas K^+ ions are more evenly distributed across all four sites. Third, at concentrations greater than about 50 mM, the occupancy and distribution of ions in the selectivity filter does not change much. The current, in contrast, continues to increase almost linearly, with a slope that is considerably steeper for K^+ than for Rb^+ .

It may seem that Rb^+ and K^+ ions permeate the K^+ channel according to different rules, but a single model can account for the data, assuming a modest difference in the way Rb^+ and K^+ ions interact with the filter. Consider all possible states of the selectivity filter in which there can exist zero, one or two ions and in which ion pairs are separated by at least a single water molecule (Fig. 4a). The connecting lines, which indicate reversible kinetic pathways, imply concerted motions of ions and water. According to the X-ray data (Figs 2 and 3), even at low millimolar concentrations of Rb^+ or K^+ , there would have to be at least one ion in the filter residing close to

one of the ends (Fig. 4a, states G and H), perhaps with about half occupancy at each end. (In an accompanying paper, we show that the state of the selectivity filter at very low ion concentration is non-conductive⁴. For both Rb^+ and K^+ , conduction begins to occur as ion entry drives the channel out of this state.) As the ion concentration is raised, by mass action the filter will more probably be found in one of the three double-occupied states (Fig. 4a, states B, C and F). Next, suppose that when K^+ is in the filter, the two double-occupied states B and C are equal in energy, and when Rb^+ is in the filter, there is an energy difference between the states such that state C is less stable than state B. In this case the electron density for K^+ would seem to be more or less evenly distributed over the filter, but for Rb^+ the density at filter positions 1 and 3 would dominate. Furthermore, to the extent that state F occurs (the double-occupied state with an ion at positions 1 and 4), we might expect in high K^+ to find the density at positions 1 and 4 to be a little stronger than at positions 2 and 3, and in high Rb^+ to find a peak at position 4 in addition to the peaks at positions 1 and 3. The data show these features (Figs 2 and 3). We therefore suggest that K^+ is very much like Rb^+ in that two ions generally occupy the filter, but in the case of K^+ , states B and C are energetically balanced, whereas in Rb^+ they are not.

The above qualitative description is reinforced by quantitative analysis through a simulation of the state diagram (Fig. 4). The general features of ion occupancy (Fig. 4c and d) and conduction (Fig. 4b) for both Rb^+ and K^+ can be reproduced. More importantly, only a single parameter of the state diagram is changed to account for the difference between the two ions: for Rb^+ there is a 5-kT energy difference between states B and C, whereas for K^+ these states are equal in energy. Either relative destabilization of state C or

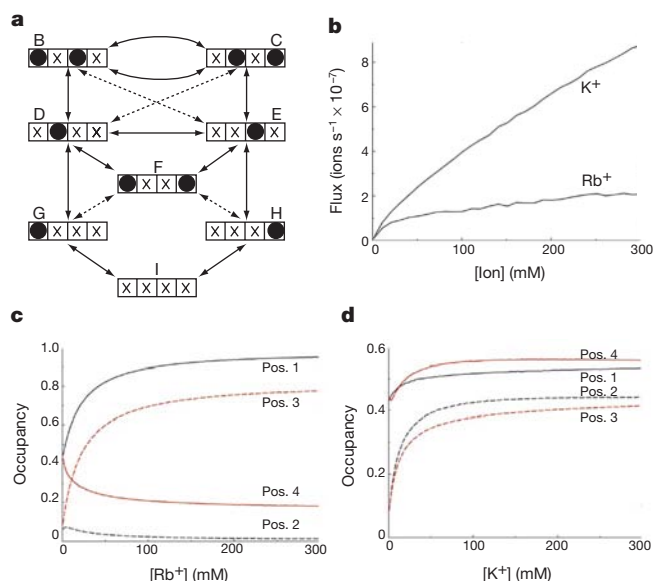


Figure 4 Analysis of a conduction-state diagram constructed on the basis of the electron density profiles. **a**, Each set of four boxes (positions 1–4 from left to right) represents a state of the selectivity filter (left side outside the cell) with a specified configuration of ions (black circles) and water (crosses). The eight states, B–I, are the configurations of 0, 1 or 2 ions, assuming that ions are separated by at least a single water molecule. Arrows show the transition paths that connect the states through concerted movements of the ion–water queue (solid lines) or by ion–water exchange at the ends of the filter (dashed lines). Two paths connecting states B and C represent a concentration-independent movement of the ions from state B (1,3 configuration) to state C (2,4 configuration) and a concentration-dependent transition due to entry of a third ion on one side, causing exit of an ion on the opposite side. **b**, Simulated flux for Rb^+ and K^+ . **c**, **d**, Mean occupancies of selectivity filter position 1, 2, 3 and 4 are shown as functions of Rb^+ (**c**) and K^+ (**d**) concentration. Values of the rate constants for the state diagram are given in the Methods.

stabilization of state B generates a similar pattern of behaviour. The unexplained, gradual increase of the peak at position 4 in the data at high Rb^+ concentrations (Fig. 2b and c) may reflect a measurable fraction of channels actually achieving a state of three Rb^+ ions in the selectivity filter (see below).

The state diagram suggests a very plausible mechanism for the main K^+ throughput cycle that occurs under physiological conditions (Fig. 4a). In the cellular environment, where the cytoplasmic K^+ concentration is always high, the double-occupied states will predominate. A more detailed picture of the cycle connecting states B and C of Fig. 4a (hereafter referred to as the 1,3 and 2,4 configurations, respectively) is shown in Fig. 5a. The essential features of this cycle are that ions are separated by a single water molecule, that a queue of ions and water move in a concerted manner, and that exchange between the 1,3 and 2,4 configurations occurs either when the ion pair ‘jumps’ between configurations (concentration-independent path) or when a third ion enters on one side causing an ion to exit from the opposite side (concentration-dependent path). At an intuitive level, this mechanism makes good energetic sense. While in the 1,3 and 2,4 configurations, each K^+ ion would reside near the centre of a box formed by eight oxygen atoms. Such coordination of single K^+ ions by eight oxygen atoms is observed in the crystal structure of the K^+ -selective antibiotic

nonactin (Fig. 5b, right)^{5,6}. The inferred water molecule located between the ions provides a second analogy to another K^+ -selective antibiotic, valinomycin (Fig. 5b, left)^{5,7}. In valinomycin the K^+ ion is coordinated in an octahedral manner, with four oxygen atoms in plane and one above and one below the ion. We imagine that during the transition between the 1,3 and 2,4 configurations a similar coordination is adopted transiently (Fig. 5a). These structural analogies to the K^+ -selective antibiotics suggest that the transition between the 1,3 and 2,4 configurations occurs without a large energy barrier.

The 1,3 and 2,4 configurations leave one ‘free’ ring of oxygen atoms on the inside and outside end of the selectivity filter, respectively (Fig. 5a). The ‘free’ ring is therefore conveniently positioned to receive an incoming third K^+ ion. Indeed, one may view the transition between the 1,3 and 2,4 configurations as the transfer of the ‘free’ oxygen ring from one end of the filter to the other, that is, a shuttling of the acceptor for the third K^+ ion back and forth across the filter. The electron density profiles in K^+ are consistent with an equal mixture of the 1,3 and 2,4 configurations and therefore the three-ion state is only implied by the cycle. (However, growth of the peak at position 4 in high Rb^+ concentrations may reflect the three-ion state reaching a detectable occupancy in that case (Fig. 2b and c)). We hypothesize that the three-ion state would be somewhat higher in energy than the double-occupied states owing to electrostatic repulsion between ions and to the necessary dehydration and rehydration of entering and exiting ions, respectively. The three-ion state could be viewed as the transition state for the throughput cycle, in which case the filter ions may move back and forth between the 1,3 and 2,4 configurations many times along the concentration-independent path before the concentration-dependent path is taken. This picture predicts a linear regime of conduction because doubling the K^+ concentration doubles the frequency of third ion entry (Figs 3a, 4b). At sufficiently high K^+ , the concentration-independent path can, in principle, become rate limiting, at which point the current should approach a maximum value. The Rb^+ data are interesting in this context. The energy difference between the 1,3 and 2,4 configurations inferred from the electron density profiles imposes a larger activation energy for going around the cycle (Fig. 6a). The lower current and shallower slope in the linear regime for Rb^+ conduction therefore agrees well with the Rb^+ electron density (Figs 2 and 4).

This conduction mechanism invokes the concerted movement of three K^+ ions in the selectivity filter (two resident ions and one entering ion) in addition to the ion located in the cavity halfway across the membrane (Fig. 1). Measurements of unidirectional ion flux in K^+ channels, and of ions interacting with blockers, have been used to estimate the number of ions seemingly present in the queue along the pore^{8–12}. These independent approaches converge on a value of 3–4 ions, consistent with the conduction mechanism

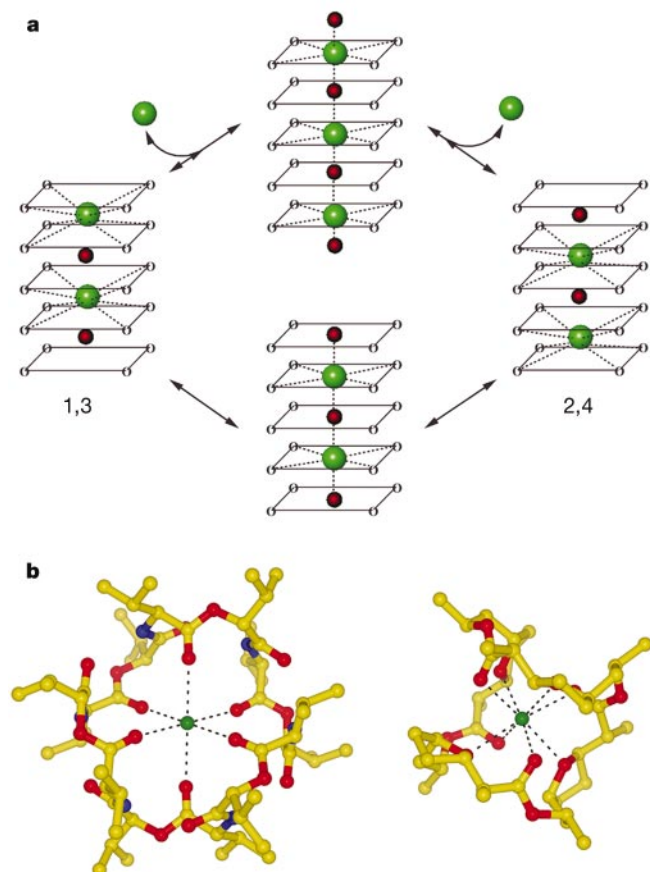


Figure 5 The biologically important throughput cycle for K^+ ions. **a**, Detailed description of the cycle connecting states B and C (Fig. 4a). The selectivity filter is depicted as five sets of four in-plane oxygen atoms (the top is outside the cell), with K^+ ions and water molecules shown as green and red spheres, respectively. K^+ ions undergo coordination by eight oxygen atoms when in the 1,3 and 2,4 configurations. Movement along either the concentration-independent path (bottom) or the concentration-dependent path (top) would involve octahedral coordination by six oxygen atoms, two provided by the intervening water molecules. **b**, The atomic structures of K^+ -selective antibiotics nonactin (right) and valinomycin (left) exhibit coordination by eight and six oxygen atoms, respectively^{5–7}.

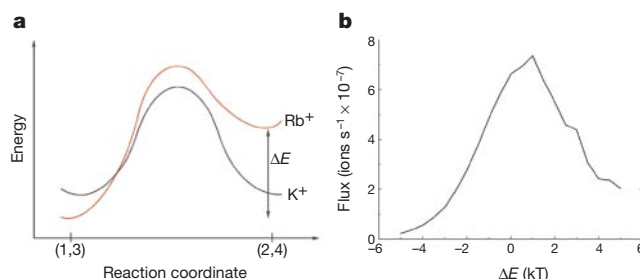


Figure 6 Energetic optimization of K^+ ion conduction. **a**, Hypothetical energy graph expresses the idea that the 1,3 and 2,4 ion configurations are energetically balanced ($\Delta E = 0$) for K^+ ions but imbalanced for Rb^+ ions. **b**, Simulated flux for the state diagram (Fig. 4a) as a function of ΔE , the energy difference between the 1,3 and 2,4 configurations. Flux corresponds to a 200 mM K^+ concentration and 10-mV driving force. Rate constants are given in the Methods.

described here. Electrophysiologists have also addressed the question of how many water molecules move in an obligatory manner across the pore with a K^+ ion. For technical reasons, this quantity is difficult to determine with great accuracy, but the number is estimated to be 1.0–1.5 water molecules per K^+ ion¹³. Although the cavity at the membrane centre contains many water molecules in it, coupling of ions and water will be confined to the narrow selectivity filter. The throughput cycle described in Fig. 5a predicts one water molecule per K^+ ion, in reasonable agreement with experiments.

Energetic optimization of K^+ conduction rate

The crystallographic and electrophysiological differences between Rb^+ and K^+ in the channel have been attributed to the relative energies of the 1,3 and 2,4 configurations for these ions. The simulation shows that the main observable differences between Rb^+ and K^+ ions can be reproduced by simply imposing an energy difference between these configurations. We wondered whether it is a coincidence that the 1,3 and 2,4 configurations are balanced energetically in the case of K^+ . There is no immediately apparent structural reason why this should be so. We propose that the observed balance is a consequence of evolutionary optimization of the rate at which ions diffuse through the pore. The basis for this proposal is simple: if the 1,3 and 2,4 configurations are imbalanced, the energy difference between them creates an inescapable energy step in the pathway around the throughput cycle. In terms of the diffusion rate, the energy step represents a kinetic barrier (Fig. 6a). Thus, by minimizing the energy difference between the 1,3 and 2,4 configurations through subtle adjustments of the protein structure, evolution has apparently minimized the kinetic barriers to achieve a high conduction rate (Fig. 6b). The channel did not evolve in response to the conduction of Rb^+ ions, the slightly different radius of which apparently creates an imbalance with kinetic consequences.

Summary

Our results lead us to conclude that, under physiological conditions, the K^+ selectivity filter usually contains two resident K^+ ions separated by a water molecule. The ion pair moves back and forth in a concerted manner between the 1,3 and 2,4 configurations until a third ion enters on one side, causing the displacement of an ion on the opposite side. The structure of the selectivity filter is designed by selection to have a maximum rate of conduction through minimization of the energy difference between the 1,3 and 2,4 ion configurations. □

Methods

Crystal preparation

KcsA was prepared as described¹. Protein was dialysed against a solution containing 5 mM lauryldodecylamine oxide (LDAO), 50 mM Tris buffer at pH 7.5, 2 mM dithiothreitol (DTT) and mixtures of NaCl, RbCl, KCl or NMG. For the growth of C2 crystals, 5 mM tetrabutylammonium chloride (TBA) was added to reduce crystal mosaicity and crystals were grown in sitting drops at 20 °C by equilibrating a 1 : 1 mixture of protein and reservoir solutions against the reservoir containing 50 mM HEPES buffer at pH 7.5, 120–300 mM $CaCl_2$, 35–50% PEG 400. Crystals were grown in the presence of 150, 300 or 500 mM KCl, 150 or 300 mM RbCl, 120 mM RbCl + 30 mM NaCl, 30 mM RbCl + 120 mM NaCl, 15 mM RbCl + 135 mM NMG and 5 mM RbCl + 145 mM NaCl. (Electron density maps from crystals containing NMG were consistent with those containing Na^+ as the ion replacing K^+ . This finding suggests that Na^+ does not enter the selectivity filter.) Crystals of space group I4 were prepared as described⁴.

Crystallographic analysis

Data were collected on frozen crystals at X-25, National Synchrotron Light Source, Brookhaven National Laboratory; at A1 and F1, Cornell High Energy Synchrotron Source; at ID19, Advanced Photon Source, Argonne National Laboratory; and at ID13, European Synchrotron Radiation Facility (Table 1). Multiple data sets were collected from crystals at each ionic condition. The data were processed and reduced with Denzo and Scalepack¹⁴. For each ionic condition, the best data set was selected based on resolution, R_{sym} and signal-to-noise ratio. Further data processing was performed with the CCP4 program suite¹⁵. All C2 crystal form data sets were scaled against the 150 mM RbCl data set and sharpened (corrected for anisotropy)¹⁶.

Model refinement of the 150 mM RbCl and 150 mM KCl structures was carried out in

Table 1 Diffraction data statistics

Salt condition	Resolution (Å)	R_{sym} * (%)	Completeness (%)
Space group C2			
RbCl 5 mM + 145 mM NaCl	3.45	10.7	86
RbCl 15 mM + 135 mM NMG	3.40	6.9	87
RbCl 30 mM + 120 mM NaCl	3.35	5.6	97
RbCl 120 mM + 30 mM NaCl	3.20	6.0	91
RbCl 150 mM	2.70	5.1	94
RbCl 300 mM	3.35	7.9	96
KCl 150 mM	2.80	7.2	93
Space group I4			
KCl 3 mM + 200 mM NaCl	2.30	5.6	99
KCl 20 mM + 180 mM NaCl	2.40	5.6	98
KCl 70 mM + 130 mM NaCl	2.20	5.4	100
KCl 200 mM	2.00	7.1	99

$$^* R_{\text{sym}} = \Sigma \Sigma (I) - I / \Sigma (I)$$

CNS with the published KcsA structure (Protein Data Bank accession code 1BL8) as an initial model¹⁷. After rigid body refinement, density modification by four-fold averaging and solvent flattening were applied before model rebuilding in program O^{15,18,19}. Using data to 2.8 Å, we refined the Rb^+ and K^+ complex structures to an R_{free} of 30.2 and 32.2, respectively. These two models, together with the original KcsA structure, were then used for determination of electron density profiles along the selectivity filter. To remove model bias, the three models, without ions, were annealed at 2000 K against their corresponding data. Rigid body and positional refinement of the three models was then performed against the data set for each ion concentration. All three models were then put into the same coordinate file and used for the calculation of difference Fourier maps at 3.45 Å. We determined one-dimensional electron density profiles by sampling the difference maps along the central filter axis of the selectivity filter using MAPMAN²⁰. To estimate the error contribution from Fourier series truncation (resolution limit), we analysed a model system of atoms in a queue with respect to the interatomic spacing and the resolution limit. We conclude that given the separation between ions in the filter, the 3.4 Å resolution limit introduces only relatively small errors in the magnitude of peaks and wells and allows us to define the ionic positions with accuracy.

All I4 crystal form data were scaled against the 200 mM KCl data set⁴. For the four K^+ concentrations, models were refined without ions or water in the pore to an R_{free} in the range 24.2–26.2%; these models were used in the calculation of difference Fourier maps at 2.40 Å. To justify the comparison of electron density maps from the I4 and C2 crystal forms, especially as the C2 form contained a TBA cation in the cavity replacing the alkali metal cation normally present, several control data sets with K^+ in the C2 form (50–500 mM), and one with Rb^+ in the I4 form (200 mM) were studied. The electron density maps showed that the different crystal forms did not affect the distribution of ions in the selectivity filter, and that the comparison of maps is valid (not shown).

Electrophysiological analysis

KcsA channels containing a six-histidine sequence on the carboxy terminus were purified and reconstituted into lipid vesicles as described²¹. Planar lipid bilayers of 1-palmitoyl-2-oleoyl phosphatidylethanolamine (POPE) (15 mg ml⁻¹) and phosphatidyl glycerol (POPG) (Avanti Polar Lipids) (5 mg ml⁻¹) in decane were painted over a 300-μm hole in a polystyrene partition separating internal and external solutions. To induce fusion of channel-containing vesicles, the external solution contained KCl or RbCl (150 mM) and HEPES (10 mM at pH 7.0) and the internal solution contained KCl or RbCl (20 mM) and succinate (10 mM at pH 4.0)^{22,23}. Following channel fusion, the internal and external solutions were adjusted by solution addition or perfusion to achieve equal concentrations of KCl or RbCl. Membrane voltage was controlled and current recorded by an Axopatch 200B amplifier with a Digidata 1322A analogue-to-digital converter and Axoclamp software (Axon Instruments).

Simulation of ion conduction

Ion conduction properties and ion probability distributions of the state diagram were analysed by a simulation in which a random variable with an exponential distribution was generated for each state²⁴. Rate constants were expressed as an exponential function of energy²⁵. Voltage dependence was applied to the rate constants by assuming that each ion crosses 0.2 of the membrane voltage difference when moving from one site to the next and when entering or exiting the filter, and that the energy barrier is located halfway between the sites. Values of the rate constants used in the simulation for K^+ were (in units of M⁻¹ s⁻¹): $k_{\text{BA}}, k_{\text{CA}}, 4 \times 10^2$; $k_{\text{DB}}, k_{\text{EC}}, 1.6 \times 10^{10}$; $k_{\text{DB}}, k_{\text{EB}}, 4 \times 10^{10}$; $k_{\text{GE}}, k_{\text{HE}}, 2 \times 10^9$; (in units of s⁻¹) $k_{\text{AB}}, k_{\text{AC}}, 1 \times 10^{11}$; $k_{\text{BC}}, k_{\text{CB}}, 2 \times 10^9$; $k_{\text{BD}}, k_{\text{CE}}, 5 \times 10^7$; $k_{\text{DE}}, k_{\text{ED}}, 1 \times 10^9$; $k_{\text{DG}}, k_{\text{EH}}, 1 \times 10^7$; $k_{\text{FE}}, k_{\text{FH}}, 5 \times 10^6$; $k_{\text{FG}}, k_{\text{HH}}, 1 \times 10^8$; $k_{\text{GD}}, k_{\text{HE}}, 1 \times 10^6$. The subscripts refer to the states in Fig. 4a, with state A representing the three-ion state connecting states B and C along the concentration-dependent path (Fig. 5a). The empty state was excluded from the simulation because it is not observed over the ion concentration range studied, and the pathways connecting states B to E and C to D were excluded for simplicity. For the Rb^+ simulation, the network was modified by adjusting the value of a single parameter corresponding to the addition of 5 kT of energy to state C relative to its neighbouring states, maintaining detailed balance in the absence of a driving force. The simulation was carried out with a 10-mV driving force.

Received 2 August; accepted 10 September 2001.

- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Eisenman, G., Latorre, R. & Miller, C. Multi-ion conduction and selectivity in the high-conductance Ca²⁺-activated K⁺ channel from skeletal muscle. *Biophys. J.* **50**, 1025–1034 (1986).
- LeMasurier, M., Heginbotham, L. & Miller, C. KcsA: It's a potassium channel. *J. Gen. Physiol.* **118**, 303–314 (2001).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Dunitz, J. D. & Dobler, M. in *Biological Aspects of Inorganic Chemistry* (eds Addison, A. W., Cullen, W. R., Dolphin, D. & James, B. R.) 113–140 (Wiley, New York, 1977).
- Dobler, v. M., Dunitz, J. D. & Kilbourn, B. T. Die struktur des KNCS-Komplexes von nonactin. *Helv. Chim. Acta* **52**, 2573–2583 (1969).
- Neupert-Laves, K. & Dobler, M. The crystal structure of a K⁺ complex of valinomycin. *Helv. Chim. Acta* **58**, 432–442 (1975).
- Hodgkin, A. L. & Keynes, R. D. The potassium permeability of a giant nerve fibre. *J. Physiol. (Lond.)* **128**, 61–88 (1955).
- Begenisich, T. & De Weer, P. Potassium flux ratio in voltage-clamped squid giant axons. *J. Gen. Physiol.* **76**, 83–98 (1980).
- Bestergaard-Bogind, B., Stampe, P. & Christophersen, P. Single-file diffusion through the Ca²⁺-activated K⁺ channel of human red cells. *J. Membr. Biol.* **88**, 67–75 (1985).
- Neyton, J. & Miller, C. Potassium blocks barium permeation through a calcium-activated potassium channel. *J. Gen. Physiol.* **92**, 549–567 (1988).
- Neyton, J. & Miller, C. Discrete Ba²⁺ block as a probe of ion occupancy and pore structure in the high-conductance Ca²⁺-activated K⁺ channel. *J. Gen. Physiol.* **92**, 569–586 (1988).
- Alcayaga, C., Cecchi, X., Alvarez, O. & Latorre, R. Streaming potential measurements in Ca²⁺-activated K⁺ channels from skeletal and smooth muscle. *Biophys. J.* **55**, 367–371 (1989).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project, No. 4. The CCP4 Suite: Programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Gamblin, S. J., Rodgers, D. W. & Stehle, T. *Proc. CCP4 Study Weekend* 163–169 (Daresbury Laboratory, Daresbury, 1996).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kleywegt, G. J. & Jones, T. A. in *From the First Map to Final Model. Proc. CCP4 Study Weekend* (eds Bailey, S., Hubbard, R. & Waller, D.) 59–66 (Daresbury Laboratory, Daresbury, 1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Kleywegt, G. J. & Jones, T. A. xdlMAPMAN and xdlDATAMAN—Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D* **52**, 826–828 (1996).
- Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single streptomyces lividans K⁺ channels. Functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).
- Heginbotham, L., Kolmakova-Partensky, L. & Miller, C. Functional reconstitution of a prokaryotic K⁺ channel. *J. Gen. Physiol.* **111**, 741–749 (1998).
- Cuello, L. G., Romero, J. G., Cortes, D. M. & Perozo, E. pH-dependent gating in the *Streptomyces lividans* K⁺ channel. *Biochemistry* **37**, 3229–3236 (1988).
- Koonin, S. E. *Computational Physics* (Benjamin/Cummings, Menlo Park, 1986).
- Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Bacon, D. & Anderson, W. F. A fast algorithm for rendering space filling molecule pictures. *J. Mol. Graph* **6**, 219–220 (1988).

Acknowledgements

We thank, for assistance, the staff at the National Synchrotron Light Source X-25; at Cornell High Energy Synchrotron Source, A1 and F1; at the Advanced Photon Source, ID19; and at the European Synchrotron Radiation Source, ID13. We thank Y. Jiang, R. Dutzler and A. Pico for assistance in data collection; B. Roux for discussions; and F. Valiyaveetil for discussion and advice on the manuscript. This work was supported by a grant from the National Institutes of Health to R.M. R.M. is an investigator in the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.M. (e-mail: mackinn@rockvax.rockefeller.edu). Coordinates have been deposited with the Protein Data Bank under accession codes 1JVM, 1K4C and 1K4D.

Chemistry of ion coordination and hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution

Yufeng Zhou, João H. Morais-Cabral*, Amelia Kaufman & Roderick MacKinnon

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Ion transport proteins must remove an ion's hydration shell to coordinate the ion selectively on the basis of its size and charge. To discover how the K⁺ channel solves this fundamental aspect of ion conduction, we solved the structure of the KcsA K⁺ channel in complex with a monoclonal Fab antibody fragment at 2.0 Å resolution. Here we show how the K⁺ channel displaces water molecules around an ion at its extracellular entryway, and how it holds a K⁺ ion in a square antiprism of water molecules in a cavity near its intracellular entryway. Carbonyl oxygen atoms within the selectivity filter form a very similar square antiprism around each K⁺ binding site, as if to mimic the waters of hydration. The selectivity filter changes its ion coordination structure in low K⁺ solutions. This structural change is crucial to the operation of the selectivity filter in the cellular context, where the K⁺ ion concentration near the selectivity filter varies in response to channel gating.

Potassium channels control the electric potential across cell membranes by catalysing the rapid, selective diffusion of K⁺ ions down their electrochemical gradient¹. The structure of the K⁺ channel has provided a firm basis for understanding the mechanisms of rapid K⁺ ion transport underlying electrical signalling in cells². Through the interactions of dehydrated K⁺ ions within the channel's selectivity filter, high conduction rates are achieved in the setting of exquisite ion selectivity³. Two fundamental questions surrounding this process are addressed in the present study. The first is how the K⁺ channel mediates the transfer of a K⁺ ion from its hydrated state in solution to its dehydrated state in the selectivity filter. The issue of dehydration is relevant to all mechanisms of selective ion transport, and because dehydration in the wrong environment is energetically costly, we should expect to discover in the K⁺ channel a very precise set of mechanisms designed to handle hydrated K⁺ ions, and to mediate their dehydration.

The second question addressed in this study is related to the cellular environment in which K⁺ channels operate: inside the cell the K⁺ concentration is greater than 100 mM, whereas on the outside the K⁺ concentration is usually less than 5 mM. The K⁺ channel gate, or door that opens and closes the pore, is located between the selectivity filter and the intracellular solution^{2,4–6}. Therefore, when the gate is open, the filter is exposed to a high K⁺ concentration from inside the cell, and when it is closed the filter is exposed to a low K⁺ concentration from outside. This is an interesting situation when one considers the structure of the selectivity filter, and the mechanism by which it conducts K⁺ ions. The filter points a large number of carbonyl oxygen atoms into the pore. Owing to the partial negative charge on these atoms, this would be an unlikely structure if it were not for the presence of dehydrated K⁺ ions in the filter. In other words, the K⁺ ions that go through the filter are actually counter-charges, necessary for its structure. The question that then arises is what happens when the channel's gate closes and the selectivity filter is in equilibrium with a low extracellular K⁺ environment. We answer this by describing the

structure of the K⁺ selectivity filter in the presence of a low K⁺ ion concentration.

K⁺ channel–Fab complex

To address the above questions it was necessary to solve the K⁺ channel structure at a resolution that would reveal ordered water molecules and protein chemistry with high accuracy. To achieve this end, we raised monoclonal antibodies against the KcsA K⁺ channel and selected clones on the basis of their ability to recognize the tetrameric but not the monomeric form of the channel^{7,8}. A K⁺ channel–Fab complex with a stoichiometry of one Fab fragment per channel subunit was produced and crystallized in space group *I*₄, with one channel subunit and Fab fragment per asymmetric unit. Frozen crystals diffracted X-rays to 2.0 Å Bragg spacings at the synchrotron. We solved phases by molecular replacement using a published Fab structure (Protein Data Bank (PDB) code 1MLC)⁹, and could easily interpret the resulting electron density map. The published KcsA K⁺ channel structure (PDB code 1BL8) was placed into the density map², followed by several cycles of rebuilding and refinement. The final model, referred to as the high-K⁺ structure (200 mM K⁺), is refined with good stereochemistry to an *R*_f and *R*_w of 23.3% and 21.8%, respectively, and contains 534 amino acids, 7 K⁺ ions, 469 water molecules and 2 partial lipids. A second structure, the low-K⁺ structure (3 mM K⁺), was solved at 2.3 Å resolution to an *R*_f and *R*_w of 23.5% and 21.8%, respectively, and contains 534 amino acids, 2 K⁺ ions, 1 Na⁺ ion, 266 water molecules and 2 partial lipids (Table 1).

The Fab fragment is attached to the K⁺ channel turret on the extracellular face of the channel (Fig. 1). All of the protein contacts within the crystal are formed between neighbouring Fab fragments, with the K⁺ channel conveniently suspended so that its detergent micelle is not involved in crystal contacts. This packing arrangement undoubtedly accounts for the high-quality X-ray diffraction^{7,8}. Furthermore, Fab attachment to the turrets leaves open a wide passageway outside the pore, so that ion binding should be unperturbed by the presence of the Fab fragments.

An electron density map surrounding the channel's selectivity filter in the high K⁺ structure is shown in Fig. 2. Density is present at four K⁺ ion-binding sites inside the selectivity filter, corresponding to positions 1–4 from outside the cell to inside. Strong electron

* Present address: Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520, USA.

Table 1 Crystallographic analysis

Data set	Data collected					
	Resolution (Å)	Redundancy, overall/outer	Completeness, overall/outer (%)	R_{sym}^* , overall/outer	I/σ , overall/outer	Reflections with $I/\sigma > 2$, overall/outer (%)
High K^+	50–2.0	3.8/3.0	99.1/95.2	0.071/0.530	22.4/2.3	79/45
Low K^+	30–2.3	4.2/2.7	99.2/95.3	0.056/0.370	24.5/2.2	80/40
	Refinement			Root mean square difference		
	Resolution (Å)	$R_w/R_{\text{f}}^\dagger$ (%)	Mean B -factor (Å ²)	Bond lengths (Å)	Bond angles (°)	B -factor (Å ²)
High K^+	40–2.0	21.8/23.3	37.4	0.007	1.4	0.7
Low K^+	30–2.3	21.8/23.5	49.2	0.007	1.4	0.7

* $R_{\text{sym}} = \sum \sum (|I| - |I|) / \sum (I)$

† $R_w = \sum |F_o - F_c| / \sum F_o$. R_{f} is defined as R_w calculated with 10% of reflections excluded from the refinement.

density is also visible at both entryways to the filter, curiously suspended along the pore axis.

K^+ hydration in the central cavity

A unique feature of the K^+ channel structure is that its ion conduction pore dilates to about 10 Å diameter at one point.

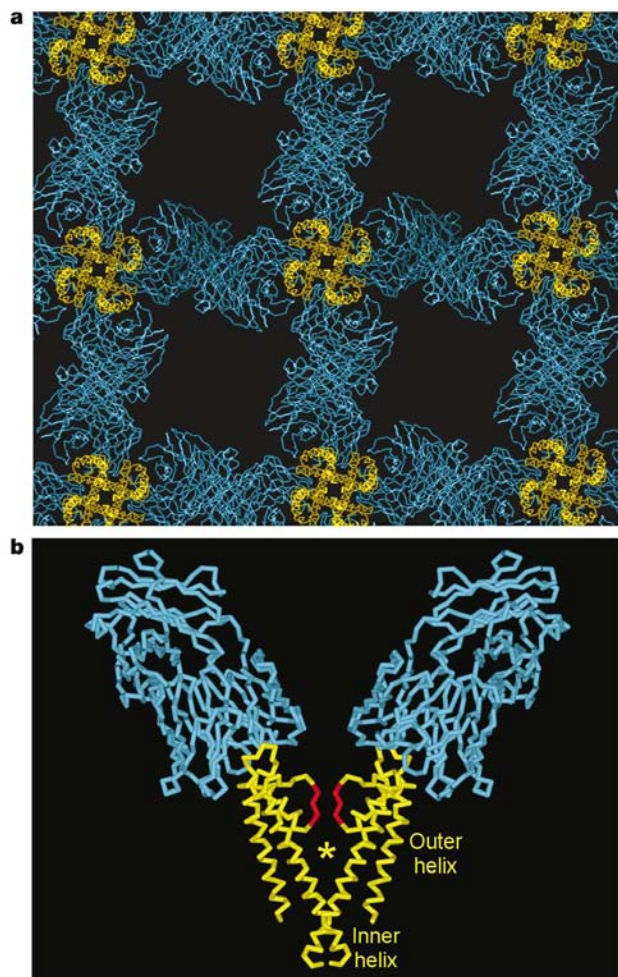


Figure 1 Fab attachment and crystal packing. KcsA (yellow) was crystallized as a complex with an antibody Fab fragment (blue). One Fab fragment is bound to the extracellular-facing turret on each K^+ channel subunit. **a**, View down the four-fold crystallographic axis of the $1/4$ cell, which corresponds to the molecular four-fold axis of the K^+ channel. **b**, Two subunits and attached Fab fragments viewed perpendicular to the four-fold axis. The transmembrane outer and inner helices are labelled. The asterisk denotes the location of the central cavity, below the selectivity filter (red).

dilation, called the central cavity, occurs at the intracellular side of the selectivity filter, halfway across the membrane (Fig. 1b). The presence of the cavity can be understood intuitively as one of the channel's mechanisms for overcoming the dielectric barrier, or repulsion by the low-dielectric membrane, by keeping the K^+ ion in a watery, high-dielectric environment^{2,10}. The electron density at the intracellular entryway to the selectivity filter corresponds to a K^+ ion at the centre of the cavity (Fig. 2, bottom density). Remarkably, this ion is fully hydrated by eight discrete water molecules, four 'above' and four 'below' the K^+ ion (Fig. 3). The appearance of water density in the cavity is influenced by the ionic species and by adding inhibitors that bind in the cavity (not shown). These observations ensure that the density surrounding K^+ in the cavity represents ordered water molecules and not a build-up of noise due to crystallographic symmetry. This structure, a K^+ hydration complex, is suspended at the cavity centre, too far from the cavity walls to make direct contact with protein atoms. There is ample space for additional water molecules in the cavity, but they are disordered: only the water molecules in the inner hydration shell of the K^+ ion are sufficiently ordered to be visible in the X-ray structure at the contour shown.

The absence of visible water molecules at the surface of the cavity is as significant as the presence of water around the K^+ ion. The cavity in KcsA and most other K^+ channels is lined mainly by hydrophobic amino acids from the inner helix (Ile 100, Phe 103; Fig. 3) that do not provide strong hydrogen-bonding donor or acceptor groups. Consequently, water in the cavity is available to interact strongly with the K^+ ion without competition from the protein surface. That said, the compelling question arises of why the water molecules around the K^+ ion are so precisely ordered, rather than being spread evenly over a spherical shell. Inspection of the cavity wall shows that the order is probably imposed by a sum of very weak, indirect hydrogen-bonding interactions mediated by certain chemical groups, such as the hydroxyl group of Thr 107, and perhaps the carbonyl oxygen atoms from the pore and inner helices. In support of this idea, electron density maps at lower contours show a less ordered water molecule acting as a bridge connecting the Thr 107 hydroxyl group to the K^+ hydration complex (not shown). The probable significance of a geometric match between the cavity and a K^+ ion is implied by the absence of a complete shell of visible water molecules when Na^+ rather than K^+ is in the cavity. Thus, a K^+ hydration complex is held in a surprisingly precise configuration at the centre of the cavity. The four-fold distribution of water molecules surrounding the K^+ ion causes one to wonder whether the fundamental structure of a hydrated K^+ ion gave rise to a four-fold symmetric channel.

Visualization of an organized K^+ hydration complex brings our understanding of the K^+ channel cavity to a deeper level. Continuum electrostatic calculations showed that a high-dielectric (water-filled) environment and oriented pore α -helices (KcsA

residues 61–73) account for K^+ ion stabilization in the cavity¹⁰. However, these electrostatic considerations predict a relatively broad energy minimum; they do not account for the very focal nature of the ion observed in electron density maps. Localization of the K^+ ion at the cavity centre is now explained by the water structure around the ion: because the cavity holds the ion and its surrounding water molecules, the ion stays at the centre. But the ordering of water molecules should in no way impede the movement of a K^+ ion through the pore, because the exchange of crystallographically ordered water molecules can occur in less than a nanosecond timescale, whereas conduction occurs on the nanosecond timescale¹¹. Through its design, the cavity ultimately achieves a very high effective K^+ concentration (~ 2 M) at the centre of the membrane, with the K^+ ion positioned on the pore axis, ready to enter into the selectivity filter.

K^+ dehydration at the extracellular entryway

The selectivity filter opens directly to the extracellular solution in such a manner as to expose four carbonyl oxygen atoms from the Gly 79 residues (Fig. 4a). These carbonyls are directed straight out into solution in a ring surrounding the perimeter of the pore entryway. Buried just beneath the protein surface, carboxyl-carboxylate pairs formed by the side chains of Glu 71 and Asp 80 provide four negative charges near the entryway (Fig. 2)¹². Thus, the extracellular pore entryway is quite electronegative and should be attractive to a cation. Electron density maps show a complex structure close to the pore entryway consisting of two strong peaks on the pore axis associated with weaker surrounding density (Fig. 4a). For electrostatic reasons it is unlikely that these two peaks represent two K^+ ions coexisting at the pore entryway. By replacing Cl^- with electron-dense I^- in the crystals, we excluded the possibility that one of the peaks is related to a K^+ counter ion (not shown). We

propose that the two on-axis peaks represent alternate positions of a single K^+ ion. The alternate positions are entirely consistent with our description of the K^+ conduction mechanism, in which the selectivity filter exists mainly in two ion configurations, K^+ -water- K^+ -water (1,3 configuration) and water- K^+ -water- K^+ (2,4 configuration) (Fig. 4b)³. Through electrostatic interactions between an ion at the entryway and ions within the selectivity filter, the occurrence of the 1,3 and 2,4 configurations should result in displacement of the entryway ion, yielding two peaks of electron density. When the entryway ion is furthest from the pore (associated with the 1,3 configuration), it must be fully hydrated, but when it moves into its position near the pore (associated with the 2,4 configuration), the carbonyl oxygen atoms from the four Gly 79 residues are close enough to displace half of the ion's hydration shell. We interpret the electron density surrounding the two on-axis peaks as a superposition of water molecules associated with the entryway K^+ ion in its two positions. It is evident that the role of the Gly 79 carbonyl oxygen atoms, pointing straight into solution, is to assist in the hydration and dehydration of a K^+ ion at the extracellular entryway.

Two filter structures under high and low K^+ concentrations

The high-resolution structure of the selectivity filter in the presence of 200 mM K^+ is shown in Fig. 5a. The K^+ ions in positions 1–4 represent the average superposition of K^+ in the 1,3 and 2,4 configurations³. The arrangement of eight oxygen atoms surrounding each K^+ ion in the selectivity filter resembles the arrangement of water oxygen atoms around K^+ in the cavity (Figs 3 and 5a). In each case, K^+ resides near the centre of a distorted cube—a square antiprism—in which the square plane of oxygen atoms ‘above’ the ion is rotated with respect to the plane ‘below’. In the filter, the oxygen– K^+ coordination distances range from 2.70 to 3.08 Å, with a

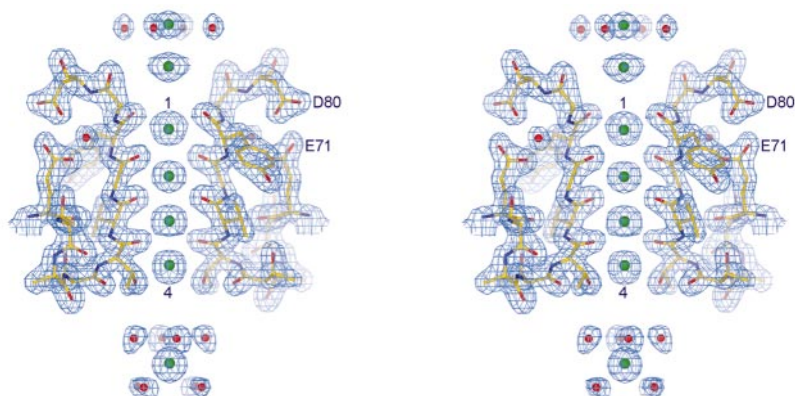


Figure 2 Stereo view of electron density in the region of the K^+ channel selectivity filter. The $2F_o - F_c$ electron density map (contoured at 2σ) covers amino acids forming the

selectivity filter (two diagonally opposed subunits are shown), with K^+ ions (green spheres) along the ion pathway, and water molecules (red spheres) in the vicinity.

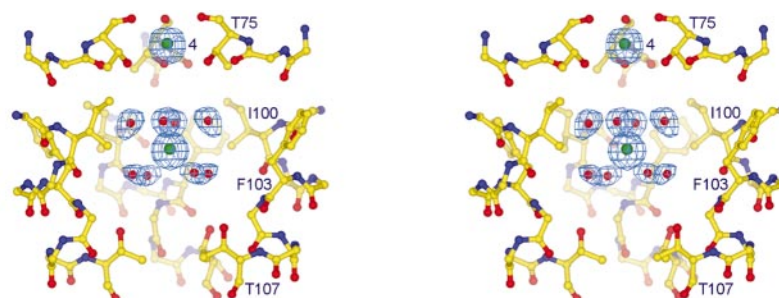


Figure 3 Stereo view of a hydrated K^+ ion in the central cavity. Eight water molecules (red spheres) surround a single K^+ ion (green sphere) in the cavity. The $2F_o - F_c$ electron density map is contoured at 2σ . Residues forming the cavity are shown in ball-and-stick

representation. For clarity, only backbone atoms and the side chains facing the cavity (Thr 75, Ile 100, Phe 103, Gly 104 and Thr 107) are shown. The subunit closest to the viewer has been removed.

mean value of 2.85 Å. The distances are very similar to those observed in the K⁺-selective antibiotics nonactin (2.73–2.88 Å) and valinomycin (2.74–2.85 Å)^{13–15}. The selectivity filter is stabilized by a network of hydrogen bonds to the amide nitrogen atoms that point away from the pore, into the protein core. The network includes a short hydrogen bond (2.65 Å) between the carboxylic group of Glu 71 and that of Asp 80 (a carboxyl-carboxylate), and a buried water molecule bonded to the amide nitrogen of Gly 79 (ref. 12). The atomic temperature factors in this region of the protein are low (15–20 Å², compared with 37 Å² overall), indicating that the filter has a fixed, well defined structure, as expected for a high-throughput catalytic device¹⁶. Certainly the filter has to adjust its structure as ions move between the 1,3 and 2,4 ion configurations³. But refinement of the two configurations indicates that the structural adjustments are very small (not shown).

The low-K⁺ structure of the selectivity filter (3 mM K⁺) is significantly different than the high-K⁺ structure (200 mM K⁺) (Fig. 5b and c). In particular, ions are absent at positions 2 and 3 and the Val 76 and Gly 77 residues have adopted completely new conformations. The Val 76 carbonyl, instead of pointing toward the pore to coordinate a K⁺ ion, is hydrogen bonded to a water molecule outside the pore, which in turn is hydrogen bonded to the amide nitrogen of Gly 77 on an adjacent subunit. The α-carbon of Gly 77 is twisted inward and occludes the pore, giving the filter an hourglass appearance. The hydrogen bond network is reorganized and a belt of buried water molecules now surrounds the perimeter of the selectivity filter inside the protein core.

This low-K⁺ structure undoubtedly represents a non-conductive state of the selectivity filter, because it is essentially pinched shut. Given the large conformational differences between the high- and low-K⁺ structures, it is likely that their rate of interconversion occurs more on the timescale of gating (milliseconds) than that of ion conduction (nanoseconds). The low-K⁺ structure underlies the electron density profiles in low concentrations of K⁺ and Rb⁺ in an

accompanying paper³, where electron density peaks for ions were observed in positions 1 and 4, with a weaker peak presumably due to a water molecule at position 3. We propose that the K⁺ channel begins to conduct ions only once the filter snaps into a main chain conformation similar to that observed in the high-K⁺ structure (Fig. 5a). The discovery of two distinct structures under different ionic conditions does not imply that the selectivity filter is poorly structured, but rather that it can adopt two unique conformations driven by the K⁺ concentration. Certainly, in high-K⁺ ion concentrations under which conduction occurs, the structure of the filter is quite inflexible.

The low-K⁺ structure explains how the selectivity filter maintains its stability in a low-K⁺ environment; a dehydrated K⁺ ion is lost but compensatory structural changes occur. These structural changes would have to be very important in a cellular environment, because the activation gate in K⁺ channels, formed by the inner helices, is located between the selectivity filter and the cytoplasm (Fig. 1)^{2,4–6}. This arrangement means that when the activation gate is open the filter ‘sees’ the high cellular K⁺ concentration (>100 mM) and when it is closed the filter ‘sees’ the extracellular solution (<5 mM) (Fig. 6). The selectivity filter presumably responds by switching between the high- and low-K⁺ structures when the channel gates.

We do not mean to imply that the selectivity filter is the activation gate. The design of a K⁺ channel is such that the filter, on the extracellular side of the pore, serves the task of selective ion conduction, whereas the activation gate, on the intracellular side, opens and closes the pore. Although these functions are separate, the activation gate and the filter can obviously be coupled. One mechanism of coupling is through the K⁺ gradient across the cell membrane, as described (Figs 5 and 6), and another potential mechanism is through direct structural perturbations propagated through the protein. By either mechanism, the high- and low-K⁺ structures offer a physical explanation for phenomena that electro-

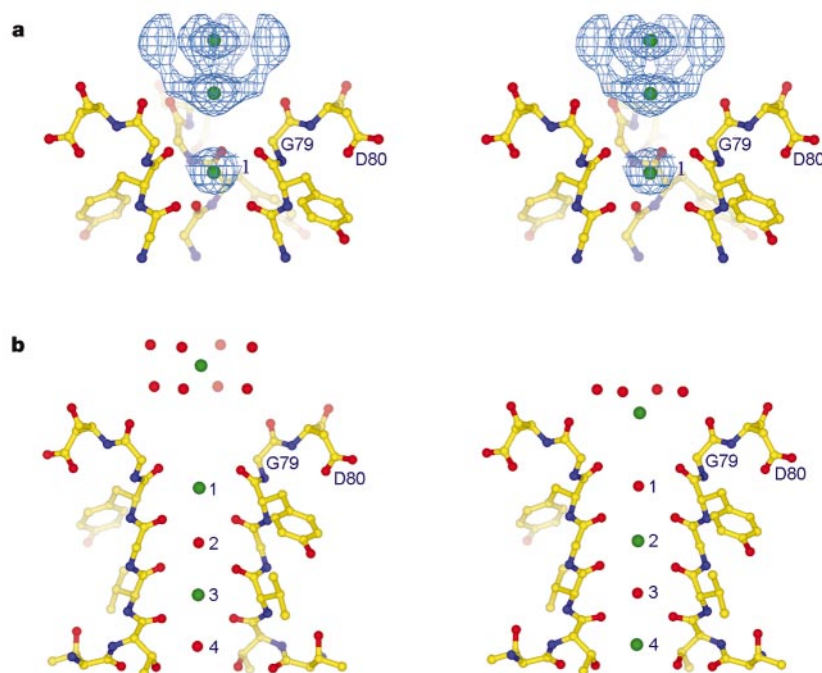


Figure 4 Potassium ion dehydration at the extracellular pore entryway. **a**, Stereo view of electron density ($2F_o - F_c$, contoured at 1σ) at the extracellular pore entryway. Channel amino acids are shown in ball-and-stick representation (with the subunit closest to the viewer removed) and K⁺ ions as green spheres. Density for the K⁺ ion at selectivity filter position 1 is shown for reference. **b**, Interpretation of the two ion peaks outside the selectivity filter. When the configuration of ions (green spheres) and water (red spheres)

inside the filter is K⁺-water-K⁺-water (left; 1,3 configuration), an ion at the entryway is displaced further away. When the configuration is water-K⁺-water-K⁺ (right; 2,4 configuration), the ion outside the filter is drawn closer to the pore. The electron density outside the filter is proposed to be the average superposition of these ion positions with their associated water molecules.

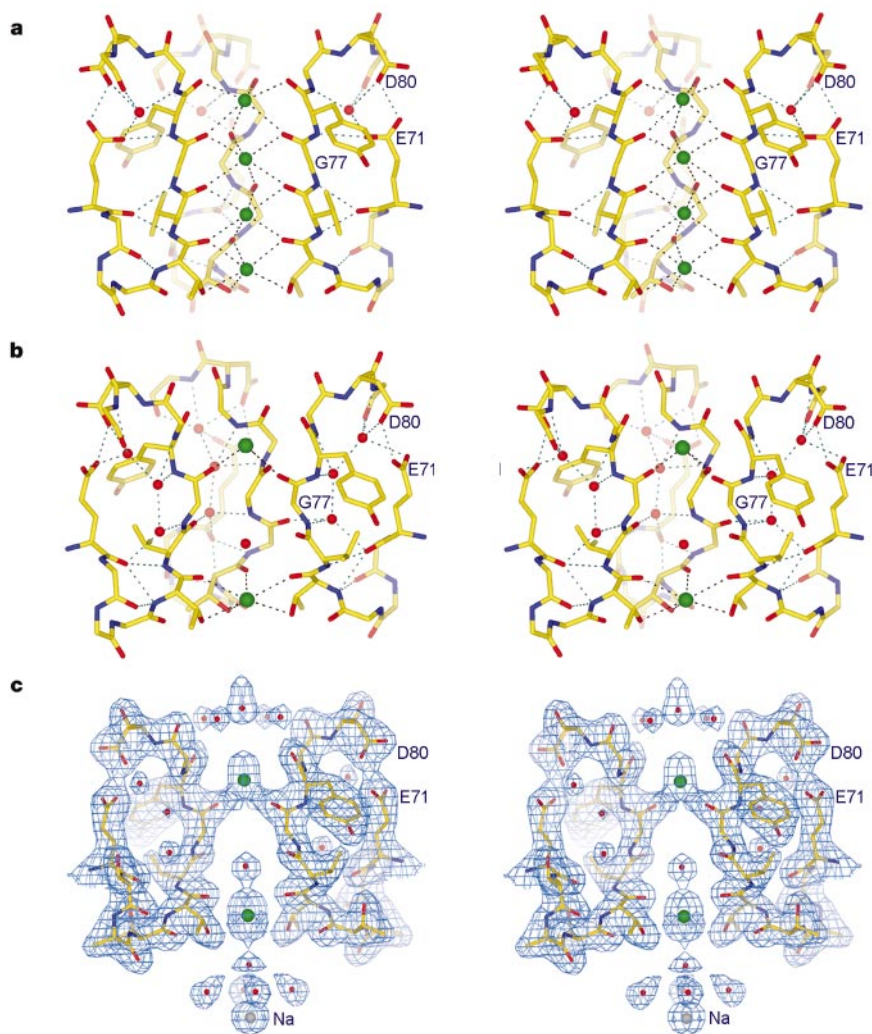


Figure 5 High- and low- K^+ structures of the selectivity filter (stereo views). **a**, The high- K^+ structure with K^+ ions (green spheres) in filter positions 1–4 (top to bottom). Each K^+ ion is located at the centre of eight oxygen atoms. The distances between the K^+ ions and the oxygen atoms (black dashed lines) above/below are: position 1, 2.80/3.08 Å; position 2, 2.72/2.83 Å; position 3, 2.85/2.70 Å; position 4, 2.94/2.88 Å. Important hydrogen bonds (blue dashed lines) and a buried water molecule (red sphere) are shown. For clarity, the

subunit closest to the viewer has been removed and side chains of the following residues are not shown: Thr 72, Ala 73, Thr 74 and Leu 81 of all three subunits, and Val 76 and Tyr 78 from the subunit at the back. **b**, The low- K^+ structure. Atoms are represented as in **a**. **c**, Electron density map ($2F_o - F_c$, 1.5σ) validating the low- K^+ structure. Density at the centre of the cavity below the filter is modelled as Na^+ (silver sphere), the predominant cation present in the crystallization solution.

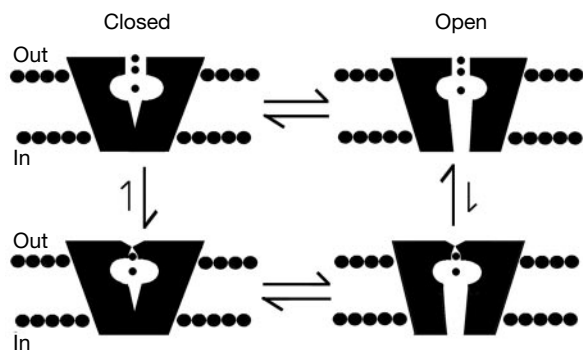


Figure 6 Biological significance of the K^+ -dependent structural change in the selectivity filter. The K^+ ion gradient across the cell membrane (high K^+ inside, low K^+ outside) causes the K^+ concentration at the selectivity filter to vary as the activation gate (near the intracellular pore opening) opens and closes. The scheme implies that when the gate is closed the filter should adopt the low- K^+ structure (lower left), and when the gate opens and K^+ flows out the filter should adopt its high- K^+ structure (upper right). Transitions between the low- and high- K^+ structures in the setting of an open activation gate (upper and lower, right) should appear as gating in single channel records.

physiologists have long referred to as permeant ion effects on gating^{17–19}.

The two filter structures also may explain apparent gating transitions that seem to occur by mechanisms other than opening and closing the activation gate. For example, in Shaker K^+ channels, which have a voltage-dependent activation gate, brief, voltage-independent closures of the open channel are observed and are sensitive to mutations near the selectivity filter^{20–22}. Such closures may represent millisecond-timescale fluctuations back and forth between the two conformations of the selectivity filter, as depicted by the two channels on the right side of Fig. 6. We propose that the underlying explanation for apparent gating of the selectivity filter, and for many of the permeant ion effects on gating, is that the selectivity filter has the capability of adopting two structures, one that conducts ions and one that does not. The ability to adopt two structures stems from the basic requirement of the filter to have an inbuilt mechanism for adjusting to the high- and low- K^+ concentrations that result from channel gating.

Summary

High-resolution structure determination of a K^+ channel–Fab

complex has allowed us to address the mechanisms of K^+ ion hydration and dehydration, and the response of the selectivity filter to the changing ionic environment imposed by channel gating. Three new principles of K^+ channel function are described. First, the central cavity holds a K^+ ion surrounded by eight ordered water molecules. The unique picture of a hydrated K^+ ion results from a geometric and chemical match between the cavity and the K^+ hydration complex. The cavity achieves a very high effective K^+ concentration (~ 2 M) at the membrane centre, with a K^+ ion positioned on the pore axis, ready to enter the selectivity filter. Second, the transfer of a K^+ ion between the extracellular solution (where a K^+ ion is hydrated) and the selectivity filter (where the ion is dehydrated) is mediated by a specific arrangement of carbonyl oxygen atoms that protrude into solution. A K^+ ion at the filter threshold senses the electrostatic field due to the ion distribution within the filter, and is drawn from a fully hydrated position to a position where it is half hydrated. Third, the selectivity filter can exist in two distinct conformations and the K^+ ion concentration drives the equilibrium between them. At a K^+ concentration of a few millimolar, the filter loses one of its dehydrated K^+ ions and undergoes compensatory structural changes that render it non-conductive. At higher K^+ concentrations the filter adopts a structure compatible with ion conduction. These conformational changes are crucial to the operation of the selectivity filter in the cellular context, where the K^+ concentration varies in response to channel gating. □

Methods

Generation and sequencing of the monoclonal antibody

Monoclonal antibody (mouse immunoglobulin- γ (IgG)) against KcsA was obtained as described²³. Total messenger RNA was isolated from the mouse hybridoma cells using TRIZOL reagents (Gibco BRL). The complementary DNA encoding the variable region of the antibody was obtained by reverse transcription and polymerase chain reaction (RT-PCR) by the 5' RACE (rapid amplification of cDNA ends) system (Gibco BRL). The PCR products were cloned into pUC18 (Gibco BRL) and then sequenced.

Purification of a KcsA–Fab complex

KcsA was purified in the detergent decylmaltoside (DM) as previously described². Thirty-five carboxy-terminal amino acids of KcsA were cleaved by chymotrypsin proteolysis. Mouse IgG was purified from mouse hybridoma cell culture supernatant using protein A affinity chromatography. Fab fragment of the antibody was obtained by papain proteolysis followed by Q-Sepharose chromatography. KcsA and Fab were mixed, and the complex obtained was purified on a Superdex 200 column equilibrated in 50 mM Tris buffer at pH 7.5, 150 mM KCl and 5 mM DM. For growing crystals in low K^+ concentration, the KcsA–Fab complex was dialysed against 50 mM Tris at pH 7.5, 2 mM KCl, 148 mM NaCl and 5 mM DM before we set up crystallization trials.

Crystal preparation

Crystals of space group I4 (high- K^+ structure, $a = b = 155.33$, $c = 76.27$, $\alpha = \beta = \gamma = 90^\circ$; and low- K^+ structure, $a = b = 155.29$, $c = 75.74$, $\alpha = \beta = \gamma = 90^\circ$) were grown at 20°C by the sitting-drop method. For each drop, concentrated KcsA–Fab complex ($7\text{--}15\text{ mg ml}^{-1}$) was mixed with an equal volume of reservoir solution (20–25% PEG 400, 50 mM magnesium acetate, 50 mM sodium acetate, pH 5–5.6). Cryoprotection was achieved by increasing the PEG concentration in the reservoir to 40% in 2.5% increments, 2–3 steps per day. The concentration of K^+ in the drop increases during the cryoprotection process; the final K^+ concentration was estimated to be about 200 mM and 3 mM for crystals set up in 150 mM KCl and 2 mM KCl, respectively. Here we refer to the structure solved with protein purified in 150 mM KCl as the high- K^+ structure, and the structure solved with protein purified in 2 mM KCl as the low- K^+ structure. All crystals were frozen in liquid nitrogen.

Crystallographic analysis

Data were collected at station X-25 of the National Synchrotron Light Source (Brookhaven National Laboratory) and at station A1 and F1 of Cornell High Energy Synchrotron Source. The data were processed with Denzo and Scalepack²⁴. Further data processing was performed with the CCP4 package²⁵. We solved the structure of the KcsA–Fab complex in high K^+ by molecular replacement with AMORE using a published Fab structure (PDB code 1MLC) as the search model^{19,26}. The published KcsA structure (1BL8) was then placed into the electron density². The model was refined by several cycles of manual rebuilding (performed with program O²⁷), followed by simulated annealing, minimization and individual B-factor refinement using CNS²⁸. All measured data between the resolution limits of $40.0\text{--}2.0\text{ \AA}$ were used for the refinement (except for a random 10% that was used for calculation of the R_p during which bulk solvent and anisotropic temperature factor

corrections were applied to the reflection data. The model contains 534 amino acids, 469 water molecules, 7 potassium ions and 2 partial lipid molecules.

The structure of the KcsA–Fab complex in 3 mM KCl (low- K^+ structure) was solved using the high- K^+ structure as a search model²⁶. Refinement was carried out with cycles of simulated annealing, as above, starting at 3000 K^{28} . All data between 30 and $2.3\text{ \AA}$ were used in the refinement (except for a random 10% used to calculate R_p). The model contains 534 amino acids, 266 water molecules, 2 potassium ions, 1 sodium ion and 2 partial lipid molecules. Both refined models have good geometry with one Ramachandran outlier, which is in the Fab. Figure 1a was prepared with program O²⁷; Figs 1b and 2–5 were prepared with Gl_Render^{29,30}.

Received 2 August; accepted 10 September 2001.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1992).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. Energetic optimization of ion conduction rate by the K^+ selectivity filter. *Nature* **414**, 37–42 (2001).
- Armstrong, C. M. Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axons. *J. Gen. Physiol.* **58**, 413–437 (1971).
- Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K^+ channel. *Neuron* **19**, 175–184 (1997).
- Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying K^+ -channel activation gating. *Science* **285**, 73–78 (1999).
- Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome *c* oxidase. *Nature Struct. Biol.* **2**, 842–846 (1995).
- Ostermeier, C., Harrenga, A., Ermler, U. & Michel, H. Structure at $2.7\text{ \AA}$ resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody Fv fragment. *Proc. Natl Acad. Sci. USA* **94**, 10547–10553 (1997).
- Braden, B. *et al.* Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* **243**, 767–781 (1994).
- Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K^+ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
- Otting, G., Liepinsh, E. & Wuthrich, K. Protein hydration in aqueous solution. *Science* **254**, 974–980 (1991).
- Speakman, J. C. Acid salts of carboxylic acids, crystals with some “very short” hydrogen bonds. *Struct. Bonding (Berlin)* **12**, 141–199 (1972).
- Dobler, v. M., Dunitz, J. D. & Kilbourn, B. T. Die struktur des KNCS-Komplexes von nonactin. *Helv. Chim. Acta* **52**, 2573–2583 (1969).
- Neupert-Laves, K. & Dobler, M. The crystal structure of a K^+ complex of valinomycin. *Helv. Chim. Acta* **58**, 432–442 (1975).
- Dunitz, J. D. & Dobler, M. in *Biological Aspects of Inorganic Chemistry* (eds Addison, A. W., Cullen, W. R., Dolphin, D. & James, B. R.) 113–140 (Wiley, New York, 1977).
- Jencks, W. P. *Catalysis in Chemistry and Enzymology* (Dover, New York, 1987).
- Swenson, R. P. Jr & Armstrong, C. M. K^+ channels close more slowly in the presence of external K^+ and Rb^+ . *Nature* **291**, 427–429 (1981).
- Miller, C., Latorre, R. & Reisin, I. Coupling of voltage-dependent gating and Ba^{2+} block in the high-conductance, Ca^{2+} -activated K^+ channel. *J. Gen. Physiol.* **90**, 427–449 (1987).
- Demo, S. D. & Yellen, G. Ion effects on gating of the Ca^{2+} -activated K^+ channel correlate with occupancy of the pore. *Biophys. J.* **61**, 639–648 (1992).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Shaker potassium channel gating. I: Transitions near the open state. *J. Gen. Physiol.* **103**, 249–278 (1994).
- Zheng, J. & Sigworth, F. J. Intermediate conductances during deactivation of heteromultimeric Shaker potassium channels. *J. Gen. Physiol.* **112**, 457–474 (1998).
- Lu, T. *et al.* Probing ion permeation and gating in a K^+ channel with backbone mutations in the selectivity filter. *Nature Neurosci.* **4**, 239–246 (2001).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project, No. 4. The CCP4 Suite: Programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Navaza, J. AMORE: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Bacon, D. & Anderson, W. F. A fast algorithm for rendering space filling molecule pictures. *J. Mol. Graph* **6**, 219–220 (1988).

Acknowledgements

We thank the staff at the National Synchrotron Light Source X-25 and Cornell High Energy Synchrotron Source A1 and F1 for assistance, Y. Jiang for help and advice at many stages of this project, F. Weis-Garcia and M. Nussenzweig for advice and teaching monoclonal methods, R. Dutzler for lipid topology files, and F. Valiyaveetil and J. Dunitz for discussions. This project was supported by a grant from the National Institutes of Health to R.M. R.M. is an investigator in the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.M. (email: mackinn@rockvax.rockefeller.edu). Coordinates have been deposited with the Protein Data Bank under accession codes 1K4C and 1K4D.

A single ion as a nanoscopic probe of an optical field

G. R. Guthöhrlein*, M. Keller*, K. Hayasaka†, W. Lange* & H. Walther*

* Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, 85748 Garching, Germany

† Communications Research Laboratory, 588-2 Iwaoka, Nishi-ku, Kobe 651-24, Japan

In near-field imaging, resolution beyond the diffraction limit of optical microscopy is obtained by scanning the sampling region with a probe of subwavelength size¹. In recent experiments, single molecules were used as nanoscopic probes to attain a resolution of a few tens of nanometres^{2,3}. Positional control of the molecular probe was typically achieved by embedding it in a crystal attached to a substrate on a translation stage. However, the presence of the host crystal inevitably led to a disturbance of the light field that was to be measured. Here we report a near-field probe with atomic-scale resolution—a single calcium ion in a radio-frequency trap—that causes minimal perturbation of the optical field. We measure the three-dimensional spatial structure of an optical field with a spatial resolution as high as 60 nm (determined by the residual thermal motion of the trapped ion), and scan the modes of a low-loss optical cavity over a range of up to 100 μm . The precise positioning we achieve implies a deterministic control of the coupling between ion and field. At the same time, the field and the internal states of the ion are not affected by the trapping potential. Our set-up is therefore an ideal system for performing cavity quantum electrodynamics^{4,5} experiments with a single particle.

One of the principal goals in optical imaging is to improve spatial resolution. Observations of the optical far field are limited by diffraction, and structures smaller than the wavelength of the radiation cannot therefore be resolved. This limitation has been overcome with the method of near-field imaging¹, in which a probe is brought close to the field to be measured. The resolution is then determined by the size of the probe. Initial experiments were performed using optical fibres with sharply pointed tips, but

better resolution was obtained by employing single molecules². The intensity distribution of the optical near field is mapped by detecting the fluorescent light emitted by the molecule as a function of its position. Using this technique, a resolution below 100 nm can be achieved^{6–9}. An inherent drawback of this method is the need for a host-crystal matrix or a substrate to accommodate and immobilize the single molecule. This causes absorption, diffraction, scattering and dephasing of the optical field under investigation in the vicinity of the probe molecule.

With the technique that we present here, the detrimental effects of the host crystal are avoided by using a single ion in a radio-frequency (r.f.) trap as a near-field probe. In many ways, a trapped ion is the perfect nanoscopic detector of the radiation field. Because of the absence of a medium from the interaction region, the optical field is not affected by the probe. The resolution may be adjusted by varying the strength of the confining r.f. field, and is then limited only by the size of the wavefunction of the ion in its motional ground state. In addition to the high spatial resolution, the spectral resolution of a single ion may be an additional advantage, as some of the narrowest spectral lines investigated so far have been measured with trapped ions.

In our experiment, a single $^{40}\text{Ca}^+$ ion is employed as a probe of the field. It is sensitive to radiation close to the resonance line $4^2\text{S}_{1/2} - 4^2\text{P}_{1/2}$ at a wavelength of $\lambda = 397 \text{ nm}$. The fluorescent light emitted by the ion is collected with a lens (numerical aperture 0.17) and detected with a photomultiplier tube (overall detection efficiency $\eta \approx 10^{-4}$). The observed fluorescence rate R is proportional to the local intensity of the optical field at the position r of the ion, that is, $R \propto I(r)$, provided that there is no saturation of the atomic transition. By scanning the position of the ion in the field and detecting the fluorescence rate at each point, a high-resolution map of the optical intensity distribution is obtained. It should be noted that a single ion can also probe the amplitude distribution $E(r)$ of the light field and hence measure its phase. Heterodyne detection of the fluorescent light must be used for such measurements, with the exciting laser as a local oscillator^{10,11}.

To demonstrate our method of single-ion mode mapping, we have investigated the eigenmodes of a Fabry–Perot resonator (Fig. 1). The transverse mode pattern is described by Hermite–Gauss functions with a beam waist $w_0 \approx 24 \mu\text{m}$, while in the direction of the cavity axis a standing wave builds up. In the

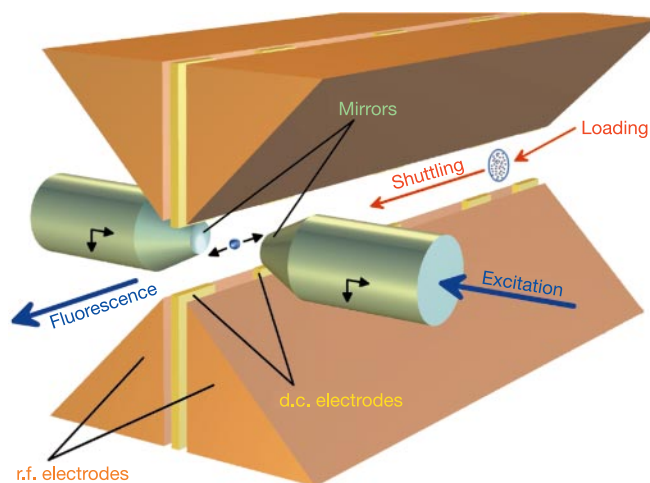


Figure 1 Experimental arrangement of trap electrodes and cavity. The ion is loaded at the rear end of the trap, and subsequently shuttled to the cavity region. Two mirrors with a radius of curvature of 10 mm and 6 mm apart form the cavity. Fluorescence is observed from the side of the cavity. For scans in the direction of the trap axis, the ion is moved with d.c. electrodes. In all other directions, the cavity is translated relative to the ion's position, as indicated by arrows.

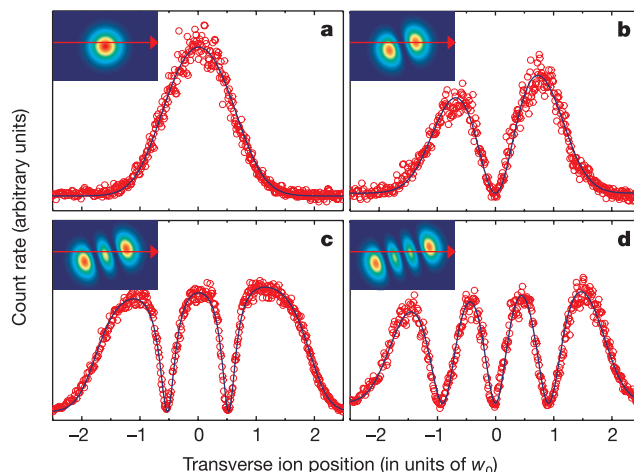


Figure 2 Transverse profiles of the Hermite–Gauss modes of the cavity, obtained by monitoring the ion's fluorescence while scanning over a range of 120 μm . The solid line is a fit including saturation of the transition. The inset shows the calculated intensity distribution of the mode and indicates the scan path. The modes are TEM_{00} (a), TEM_{01} (b), TEM_{02} (c) and TEM_{03} (d). Note the influence of saturation in c, where a slightly higher laser intensity was injected into the cavity. w_0 , beam waist.

experiment, a particular cavity mode is excited by a laser beam with a power of a few hundred nanowatts at a wavelength of 397 nm. The length of the cavity is actively stabilized to this mode.

An ion is loaded in the trap after electron-impact ionization of calcium atoms. As the electron beam and the calcium beam would degrade the optical mirrors and make stable trapping difficult, we use a linear trap and load it in a region spatially separated from the observation zone, as shown in Fig. 1. Subsequently, d.c. electrodes along the axis are employed to shuttle the ion over a distance of 25 mm to the uncontaminated end of the trap, where the cavity is located, oriented at right angles to the trap axis. Residual d.c. fields in the radial direction must be carefully compensated with correctional d.c. voltages to place the ion precisely on the nodal line of the r.f. field (coinciding with the trap axis), to avoid the trapping field exciting an oscillatory motion of the ion (micromotion).

In the direction of the trap axis, the ion is confined in a d.c. potential well, which is approximately harmonic with an oscillation frequency of $\omega_z/2\pi \approx 300$ kHz. By applying asymmetric voltages, the minimum of the potential well and thus the equilibrium position of the ion is moved along the trap axis. By simultaneously monitoring the fluorescence, we have sampled one-dimensional cross-sections of the cavity mode. The width of the ion's wavefunction in the axial potential well is a few hundred nanometres, which provides sufficient resolution to map the transverse mode pattern with an intensity distribution varying on a scale given by the cavity waist w_0 .

Figure 2 shows scans of the first four TEM_{0n} modes of the cavity obtained in this way. The indices 0 and n specify the number of nodes of electromagnetic field in the vertical and horizontal direction, respectively. The fluorescence data are not entirely symmetric because of a small displacement and rotation of the cavity eigenmodes with respect to the trap axis. In each plot, an inset indicates the path along which the ion was scanned. The solid curves in Fig. 2 are obtained from a fit using Hermite–Gauss functions, and take

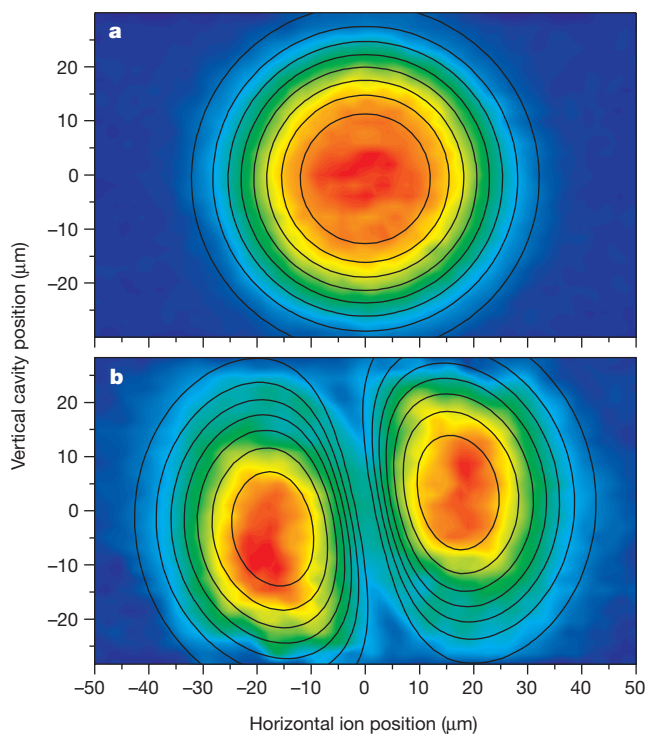


Figure 3 Two-dimensional images of the cavity field taken over an area of $100 \times 60 \mu\text{m}^2$. **a**, TEM₀₀ mode. **b**, TEM₀₁ mode. In the horizontal direction the ion was moved, while vertically the cavity position was changed relative to the ion. False colours represent the measured fluorescence count rate; the contour lines indicate the theoretical fluorescence pattern.

into account saturation of the ion's transition. In all cases, the correspondence with the measured fluorescence is excellent.

The ion's motion must be restrained to the trap axis, because off the axis the r.f. field of the trap would lead to micromotion. To scan other dimensions of the field, the sample needs to be moved. In our experiment this is done by translating the entire cavity assembly perpendicular to the trap axis, using a piezoelectric transducer. Examples of two-dimensional mode maps, scanned by combining the horizontal movement of the ion with a vertical movement of the cavity, are given in Fig. 3.

Radio-frequency confinement of the ion perpendicular to the trap axis is also harmonic, but the corresponding oscillation frequency $\omega_r/2\pi \approx 1.1$ MHz is larger than the axial frequency ω_z , so that field structures in the radial direction of the trap are better resolved. The resolution achieved with our method may be determined most accurately by probing the standing-wave field created between the cavity mirrors, which oscillates on a scale of $\lambda/2$. To this end, the cavity was moved parallel to its axis while keeping the ion stationary and monitoring its fluorescence. Figure 4 shows the mapping of the cavity field obtained in this way. A pronounced standing-wave pattern is observed with a visibility of 40%.

At present, the visibility in our experiment is given by the thermal wave packet spread of the calcium ion under conditions of Doppler cooling. In the direction of the cavity axis, cooling is provided by the cavity field itself, which is red-detuned with respect to the ion's transition for this purpose. The limiting temperature is then determined by the linewidth Γ of the atomic transition and is given by $\hbar\Gamma/2k_B \approx 500 \mu\text{K}$. The transverse degrees of freedom of the ion are cooled with an additional laser beam injected from the side of the cavity, which is switched off during the measurement. Assuming thermal equilibrium, there is a direct connection between the spread b of the ion's wavefunction, which is identical to the resolution, and the visibility V :

$$V \approx \exp[-(2\pi b/\lambda)^2] \quad (1)$$

The observed visibility corresponds to a wave packet 60 nm wide, indicating a temperature of the ion slightly above the Doppler temperature. The resolution of single-ion mode mapping in this case is almost one order of magnitude better than the wavelength of the transition.

Even higher resolution is within reach of standard ion-trapping technology. By using larger driving frequencies and higher voltages, the trap is made steeper, implying a wave packet size below 10 nm. Another factor of two is gained by cooling the ion to the ground state of motion, which has been achieved in several experiments^{12–14}. Finally, the technique of motional squeezing¹⁵ of the ion's wavefunction may be employed to create a 'breathing' wave packet which

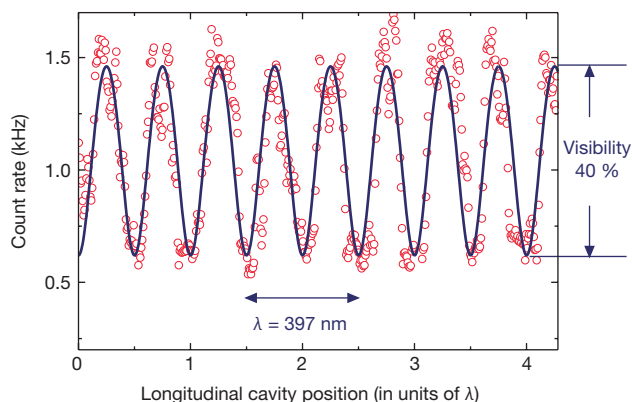


Figure 4 Single-ion mapping of the longitudinal structure of the cavity field. The visibility is determined by the residual thermal motion of the Doppler-cooled ion, and corresponds to a resolution of 60 nm. The localization of the ion's wave packet in this measurement is 16 nm.

periodically reaches a position uncertainty below 1 nm. In this case, stroboscopic detection must be used. Thus a single ion as a probe of the optical field has the potential for atomic-scale resolution, and it does not disturb the field to be measured.

An important issue, especially with regard to applications in cavity quantum electrodynamics (QED) is the question of how well the ion is localized in the optical field—that is, how accurately the centroid of the ion's wavefunction may be determined. With the spatial variation of the local fluorescence rate as the only parameter available from which to infer a change of position of the ion, the largest position sensitivity in our set-up is obtained at maximum slope of the standing wave. At this point, the localization Δx is completely determined by the signal-to-noise ratio SNR and the visibility V :

$$\Delta x \approx \frac{\lambda}{4\pi V(\text{SNR})} \quad (2)$$

With a signal-to-noise ratio of 5:1, reached with an integration time of 200 ms at each point, a localization of approximately 16 nm or $\lambda/25$ was measured. Even in the presence of noise, our measurement with a scanned single particle thus constitutes the most precise measurement to date of the spatial structure of an optical field^{6–9}. In principle, Δx could be further reduced if the signal-to-noise ratio were improved by using a longer averaging time.

As excellent spatial resolution is achieved without material support structures, a single ion is a valuable probe when minimal perturbations of the optical field are desired. However, the technique of localizing an ion in a resonator has applications beyond subwavelength imaging. The interaction of a single atom with a single field mode of a high-finesse cavity has been the subject of a number of experiments in the field of cavity QED^{4,5}. But most of these investigations have suffered from a lack of control over the position of the atom, which results in non-deterministic fluctuations of the coupling between atoms and field. In this context, the strong localization and position control, demonstrated here for a single ion in a cavity field, is an important advance, and we expect that it will be a key to future progress in the field. We plan to implement two experiments exploiting the localization of an ion in a cavity in our system. By pulsed excitation of a maximally coupled ion, single-photon wave packets may be emitted from the cavity on demand^{16,17} (single-photon gun). Under conditions of strong coupling, a single calcium ion in the cavity provides sufficient gain to build up a laser field¹⁸. Analogous to a single ion in free space, which has been shown to be an excellent source of antibunched light^{10,19}, radiation from a single-ion laser has non-classical photon statistics and correlations.

An equally attractive goal in the area of cavity QED is the simultaneous interaction of two or more ions with a single cavity mode. Owing to its linear geometry, our trap is well suited to storing multiple ions along the axis. As a first test, we have placed an array of two ions in the cavity field and observed the total fluorescence. We succeeded in matching the ion crystal to the two maxima of the TEM₀₁ mode of the cavity. In such a configuration, the cavity field may be used to entangle the two ions^{20,21}. This is a promising alternative to schemes involving the ions' motional degrees of freedom, as there is no need for cooling the vibrational modes of the ion chain below the Doppler temperature. We consider that using a cavity to perform quantum operations on adjacent pairs of ions in a long chain is a viable route to a scalable quantum computer. □

Received 9 July; accepted 7 September 2001.

- Pohl, D. W., Denk, W. & Lanz, M. Optical stethoscopy: image recording with resolution $\lambda/20$. *Appl. Phys. Lett.* **44**, 651–653 (1984).
- Betzig, E. & Chichester, R. J. Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422–1425 (1993).
- Michaelis, J., Hettich, C., Mlynec, J. & Sandoghdar, V. Optical microscopy using a single-molecule light source. *Nature* **405**, 325–328 (2000).

- Hood, C. J., Lynn, T. W., Doherty, A. C., Parkins, A. S. & Kimble, H. J. The atom-cavity microscope: Single atoms bound in orbit by single photons. *Science* **287**, 1447–1453 (2000).
- Pinkse, P. W. H., Fischer, T., Maunz, P. & Rempe, G. Trapping an atom with single photons. *Nature* **404**, 365–368 (2000).
- Michaelis, J. *et al.* A single molecule as a probe of optical intensity distribution. *Opt. Lett.* **24**, 581–583 (1999).
- Sick, B., Hecht, B., Wild, U. P. & Novotny, L. Probing confined fields with single molecules and vice versa. *J. Microsc.* **202**, 365–373 (2001).
- Gotzinger, S., Demmerer, S., Benson, O. & Sandoghdar, V. Mapping and manipulating whispering gallery modes of a microsphere resonator with a near-field probe. *J. Microsc.* **202**, 117–121 (2001).
- Veerman, J. A., Garcia-Parajo, M. F., Kuipers, L. & Van Hulst, N. F. Single molecule mapping of the optical field distribution of probes for near-field microscopy. *J. Microsc.* **194**, 477–482 (1999).
- Höfges, J. T., Baldauf, H. W., Eichler, T., Helmfrid, S. R. & Walther, H. Heterodyne measurement of the fluorescence of a single trapped ion. *Opt. Commun.* **133**, 170–174 (1997).
- Höfges, J. T., Baldauf, H. W., Lange, W. & Walther, H. Heterodyne measurement of the resonance fluorescence of a single ion. *J. Mod. Opt.* **44**, 1999–2010 (1997).
- Monroe, C. *et al.* Resolved-side-band Raman cooling of a bound atom to the 3D zero-point energy. *Phys. Rev. Lett.* **75**, 4011–4014 (1995).
- Peik, E., Abel, J., Becker, T., von Zanthier, J. & Walther, H. Sideband cooling of ions in radio-frequency traps. *Phys. Rev. A* **60**, 439–449 (1999).
- Roos, C. F. *et al.* Experimental demonstration of ground state laser cooling with electromagnetically induced transparency. *Phys. Rev. Lett.* **85**, 5547–5550 (2000).
- Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. Generation of nonclassical motional states of a trapped atom. *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
- Law, C. K. & Kimble, H. J. Deterministic generation of a bit-stream of single-photon pulses. *J. Mod. Opt.* **44**, 2067–2074 (1997).
- Hennrich, M., Legero, T., Kuhn, A. & Rempe, G. Vacuum-stimulated Raman scattering based on adiabatic passage in a high-finesse optical cavity. *Phys. Rev. Lett.* **85**, 4872–4875 (2000).
- Meyer, G. M., Briegel, H. J. & Walther, H. Ion-trap laser. *Europhys. Lett.* **37**, 317–322 (1997).
- Diedrich, F. & Walther, H. Nonclassical radiation of a single stored ion. *Phys. Rev. Lett.* **58**, 203–206 (1987).
- Pellizzari, T., Gardiner, S. A., Cirac, J. I. & Zoller, P. Decoherence, continuous observation, and quantum computing—a cavity QED model. *Phys. Rev. Lett.* **75**, 3788–3791 (1995).
- Zheng, S. B. & Guo, G. C. Efficient scheme for two-atom entanglement and quantum information processing in cavity QED. *Phys. Rev. Lett.* **85**, 2392–2395 (2000).

Acknowledgements

This work was supported by the European IST/FET programme QUBITS and by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to W.L. (e-mail: Wolfgang.Lange@mpq.mpg.de).

Ultrafast generation of magnetic fields in a Schottky diode

Y. Acremann*, M. Buess*, C. H. Back*, M. Dumm†, G. Bayreuther† & D. Pescia*

*Laboratorium für Festkörperphysik, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland

†Institute für Experimentelle und Angewandte Physik, Universität Regensburg, Universitätsstr. 31, 93040 Regensburg, Germany

For the development of future magnetic data storage technologies, the ultrafast generation of local magnetic fields is essential. Subnanosecond excitation of the magnetic state has so far been achieved by launching current pulses into micro-coils and micro-striplines^{1–6} and by using high-energy electron beams⁷. Local injection of a spin-polarized current through an all-metal junction has been proposed as an efficient method of switching magnetic elements^{8,9}, and experiments seem to confirm this^{10–13}. Spin injection has also been observed in hybrid ferromagnetic–semiconductor structures^{14,15}. Here we introduce a different scheme for the ultrafast generation of local magnetic fields in such a hybrid structure. The basis of our approach is to optically pump a Schottky diode with a focused, ~ 150 -fs laser pulse. The laser pulse generates a current across the semiconductor–metal junction, which in turn gives rise to an in-plane magnetic field. This scheme combines the localization of current injection techniques^{11–13,16} with the speed of current generation at a

Schottky barrier. Specific advantages include the ability to rapidly create local fields along any in-plane direction anywhere on the sample, the ability to scan the field over many magnetic elements and the ability to tune the magnitude of the field with the diode bias voltage.

Figure 1a shows schematically the junction used in this experiment. It consists of an n-doped GaAs(001) substrate with a 12-monolayer-thick (12-ML) $\text{Co}_{25}\text{Fe}_{75}$ film grown in ultrahigh vacuum^{17,18}. A 30-ML gold cap layer is deposited to protect the sample from corrosion, as the experiments are performed at ambient pressure. Static Kerr measurements show a square hysteresis with a coercive field of ~ 24 Oe for a field applied along the easy, [110], in-plane direction. An ohmic contact is attached to the bottom of the device, and a voltage can be applied between the top metallic film and the GaAs substrate. The ultrashort light pulses (~ 150 fs) from a mode-locked Ti:sapphire laser are used for two distinct purposes: (1) as a pump: the laser pulse with ~ 800 nm wavelength is focused to a spot of ~ 2 μm diameter (energy per pulse is 0.5 – 1.5 mJ cm^{-2}). This beam produces the magnetic field pulse—as we will explain later in detail—required to tip the magnetization vector out of its equilibrium state. (2) As a probe: after frequency doubling, the pulse is directed through the same polarization-conserving objective (spot size ~ 1 μm , energy per pulse ~ 0.75 mJ cm^{-2}).

The change of polarization of the reflected light is used to detect the perpendicular component of the magnetization via the magneto-optic Kerr effect in the polar geometry⁴. Various filters are used to protect the detector from the reflected pump beam. For detection, the pump beam is mechanically chopped at a frequency of

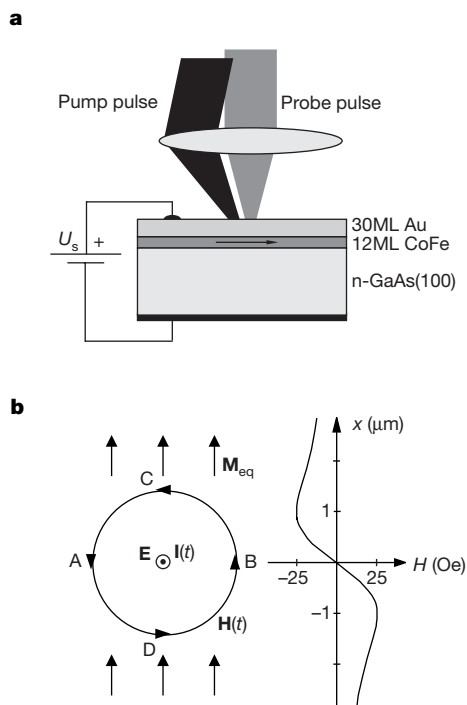


Figure 1 Diagram of the experimental set-up and the symmetry of the magnetic field. **a**, A laser pulse train with a repetition rate of 4.76 MHz and pulse duration of 150 fs is used as the clock of the experiment. Both pump and probe beam are fed into a 100 $\times$ objective lens (numerical aperture 0.95) and focused onto the Au/Fe/GaAs(001) structure. The spot of the pump beam can be moved in a circle around the probe beam. **b**, The symmetry of the resulting pulse field H in the sample plane, and the direction of the equilibrium magnetization $\mathbf{M}_{\text{eq}}$. In the reference frame of the pump beam we measure at positions A–E. Also shown is the dependence of the magnetic field amplitude on positions relative to the pump beam. At bias voltage $U_s = 0$ we estimate a maximum strength of ~ 25 Oe (see text). ML, monolayer.

1.6 kHz, and the signal is measured with a Wollaston prism followed by two photodiodes forming a bridge using lock-in techniques. To avoid overheating the optical components and the sample, we do not use the full ~ 76 -MHz laser repetition rate: this rate is reduced by a pulse picker, which works at a maximum rate of ~ 4.76 MHz. By varying the time delay between pump and probe pulse, we can measure the time evolution of the perpendicular component of the magnetization vector with picosecond accuracy⁴. In addition, the pump pulse can be misaligned by ~ 1 μm with respect to the probe pulse, so that the pump pulse can be positioned within a small circle around the probe pulse (Fig. 1a).

Assuming that the pump pulse launches a current through the junction perpendicular to the film plane, the resulting magnetic field lines should be circles in the plane of the junction centred around the pump-pulse location (Fig. 1b). The magnetization of the film is in the same plane along the easy [110]-direction (arrows in Fig. 1b): thus, along a circle, all angles between the equilibrium magnetization vector and the pulsed magnetic field vector are realized. At locations where the field is parallel or antiparallel to the magnetization, $\mathbf{M} \times \mathbf{H} = 0$, no torque acts on the equilibrium magnetization vector and no excitation takes place (locations marked A and B in Fig. 1b) (some residual magnetic excitation remains, owing to slight misalignment of the beams or a residual perpendicular magnetic field component). In contrast, at locations where the magnetization is perpendicular to the applied field pulse, the torque is maximized and deviations from the equilibrium configuration are optimally excited, leading to ferromagnetic resonance (FMR)-like oscillations (locations C and D in Fig. 1b). If the pump and the probe beam are aligned to the same point on the sample, only small oscillations of the magnetization are observed, owing to cancellation and/or decreasing strength of the magnetic field (E in Fig. 1b). On the right-hand side of Fig. 1b, we have plotted

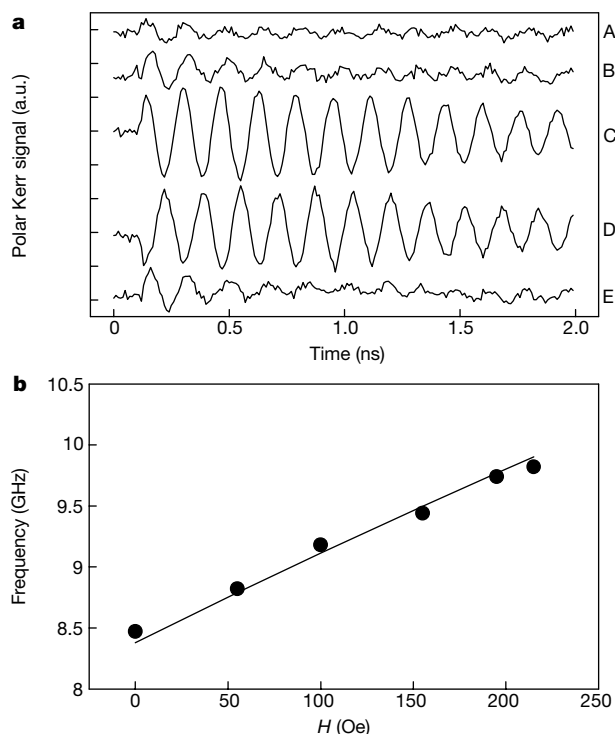


Figure 2 Experimental signature of precessional motion. **a**, Polar Kerr signal in zero applied magnetic field as a function of the delay between pump and probe pulse for the measurement positions A–E (Fig. 1b). The curves are offset for clarity (a.u., arbitrary units). Bias voltage, -1.5 V. **b**, Oscillation frequency as a function of magnetic field applied along the easy [110]-direction measured at position C. Bias voltage, -1.5 V. The solid line represents a fit to the Kittel formula (see text).

the magnetic field profile as a function of the relative position of pump and probe laser beams, calculated by inserting in Ampère's law the measured laser spot size and assuming a gaussian current profile. The strength of the field will be discussed below.

The magnetic field produced in the device shown in Fig. 1 is indeed capable of triggering precessional motion in the magnetically active side of the device. In Fig. 2a we report the oscillating perpendicular component as a function of the time delay between pump and probe pulse in the five different regions of Fig. 1b. As expected, the largest oscillating component is observed in the regions of maximum torque—see traces C and D. We note a crucial detail of traces C and D: they are out of phase by π . This phase shift mirrors the opposite sign of the torque in regions C and D. Further evidence that the traces in Fig. 2a describe precessional motion in the ferromagnetic film is provided by the dependence of the oscillation frequency measured at position C on a static magnetic field H along the easy axis (Fig. 2b). The analysis of static Kerr magnetometry reveals the existence of a large uniaxial anisotropy (anisotropy field $H_u = 570 \pm 80$ Oe). Neglecting a smaller contribution to the anisotropy energy due to a quartic component ($H_4 = 13 \pm 5$ Oe), we can use Kittel's expression for the FMR frequency¹⁹ as a fit to our data: $\omega(H) = \gamma \sqrt{(H + H_u)(H + H_u + (4\pi M_s)_{\text{eff}})}$, M_s being the saturation magnetization. The continuous line through the data points

in Fig. 2b was constructed using Kittel's equation with $(4\pi M_s)_{\text{eff}} = 15$ kOe (determined by superconducting quantum interference device (SQUID) magnetometry) and $\gamma = -\gamma_e g/2$ (γ_e being the gyromagnetic ratio of the electron) and g the Landé factor. As a consistency check, we have determined, from the fit, that the g -factor for the CoFe alloy is ~ 2.0 , which is a plausible value for a ferromagnet consisting of Fe and Co. Essentially, in the applied field regime used in Fig. 2b, the value of the precessional frequency is well accounted for by the internal fields given by the uniaxial magnetic anisotropy and the shape anisotropy. Note that reducing the intensity of the pump beam by a factor of three leaves the precessional frequency unchanged. We conclude that the excited perpendicular component is small enough to justify the use of Kittel's equation, and that the magnetic properties of the ferromagnet are not modified by the pump beam on the timescale of the precession^{20,21}.

The experimental results of Fig. 2 show the development of a sizeable ultrashort magnetic field at the Schottky junction upon local current injection. We now estimate the strength of the magnetic field. The key mechanism leading to photocurrent is the excitation of electron–hole pairs by the pump laser. The electric field in the depletion region of the Schottky barrier separates them, the holes being injected into the ferromagnet, and the electrons being swept into the semiconductor, thus producing a current. A second contribution to the transient current pulse are the electrons which are emitted from the ferromagnet into the semiconductor. This second contribution might be as important as the hole injection, and cannot be distinguished in the present experiment. The extra charge induced by the laser pulse modifies the reflectivity R : measuring R as a function of time can be used to follow the time evolution of the extra charge (Fig. 3a inset). We observe a sudden increase of $R(t)$ due to carrier creation, followed by a slower exponential decay $R(t) = R_0 e^{-t/\tau}$. As the current $I(t)$ equals the flowing charge per unit time, we can estimate the transient current from the derivative of the decaying reflectivity signal: $I(t) \propto 1/\tau e^{-t/\tau}$. The total number of carriers per pulse (determined from the macroscopic photoinduced current—in the present case, ~ 20 mA at zero bias voltage) leads to a magnetic field that rises within a few picoseconds to a peak value which we conservatively estimate, using Ampère's law, to be ~ 25 Oe at zero bias voltage and decays with a time constant related to τ . As the maximum field scales with $1/\tau$, we expect roughly a doubling of the maximum field strength when the bias voltage, U_s , is reduced from 0 to ≤ -1.5 V and a corresponding doubling of the amplitude of the oscillations. This estimate is in line with the data of Fig. 3b (see below). (We have used the Landau–Lifshitz equation to estimate the role of the exponential factor in determining the amplitude of the oscillations, and found that in the temporal range of interest (60–200 ps) the oscillation amplitude only depends on the prefactor, thus justifying our statement that the maximum field 'scales' with $1/\tau$.) Note that the linearity of the precessional motion excludes magnetization reversal. The absence of reversal is not surprising, as the fields required for precessional reversal are of the order of the anisotropy field⁷.

The magnetic field can be tuned by varying the voltage across the Schottky barrier. Figure 3a shows that the decay time τ depends on the voltage applied across the barrier. This is due to the generated carriers²² separating within the Schottky barrier, reducing the built-in electric field and thus delaying the sweep-out time. At low pumping power (~ 0.5 mJ cm⁻²) the space-charge effect should be negligible, the sweep-out being dominated by the built-in field, and no strong dependence of the sweep-out time on applied voltage is expected. Indeed (Fig. 3a, open circles), τ is almost independent of the applied voltage. In contrast, at high pumping power (1.5 mJ cm⁻²) the space-charge effect is more drastic, leading to a strong reduction of the total electric field in the depletion region for $U_s \geq -0.5$ V and a consequent strong dependence of τ on the applied voltage (filled circles in Fig. 3a)²².

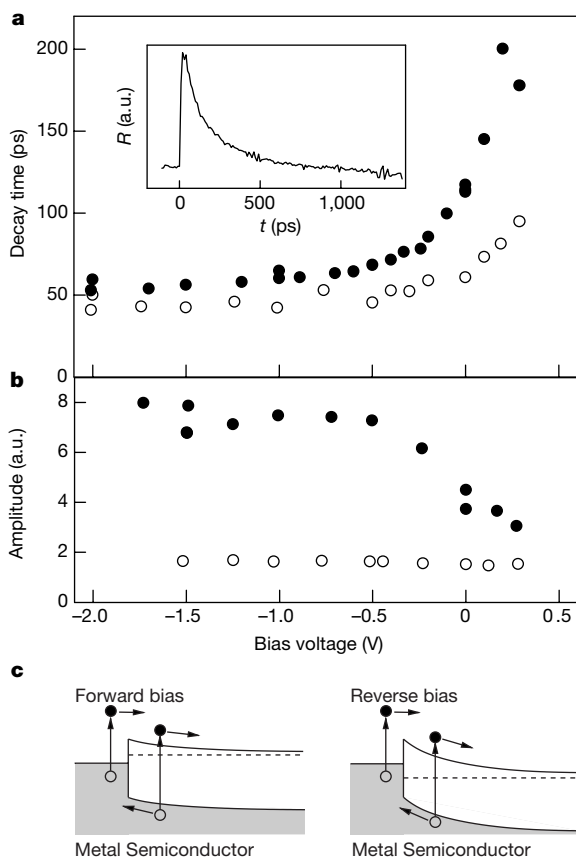


Figure 3 Voltage characteristics of the diode. **a**, The change of the reflectivity R of the sample as a function of delay between pump and probe pulse is shown in the inset. The decay was fitted with an exponential function $R = R_0 e^{-t/\tau}$. The sweep-out time τ is plotted as a function of the bias voltage for low (open symbols) and high (filled symbols) pumping power. **b**, Amplitude of the ferromagnetic-resonance-like oscillations in zero applied magnetic field as a function of the d.c. bias voltage applied across the junction. Note the different amplitudes for high (filled circles) and low (open circles) pump intensity. **c**, Band bending at the metal–semiconductor junction for positive (forward) bias (left-hand side) and for negative (reverse) bias (right-hand side). The band bending is in both ranges ‘upwards’ toward the metal.

The voltage dependence of τ produces a voltage dependence of the current, which is observed by measuring the amplitude of the FMR-like oscillations as a function of U_s (in Fig. 3b at position C). Again, we consider a high pumping power (filled circles) of 1.5 mJ cm^{-2} and a pumping power reduced by a factor of about three (open circles, 0.5 mJ cm^{-2}). In this figure there are two voltage ranges: in one range ($U_s \leq -0.6 \text{ V}$)—the reversed bias voltage range in the diode—the amplitude is almost constant, and scales roughly linearly with the pumping power. In this range, the whole system formed by the Schottky barrier and the magnetic film acts in a linear way. In the other range ($U_s \geq -0.6 \text{ V}$)—at high pumping power—the amplitude decreases until $U_s \approx 0.4 \text{ V}$ (we did not apply higher voltages to avoid damaging the diode because of the high d.c. current flow). For low pumping power the amplitude does not show a significant change, and the curves for high and low pumping power begin to approach each other for $U_s \geq -0.5 \text{ V}$. Notice that, when U_s changes sign, the oscillation neither reverses its sign nor vanishes. Evidently, we are dealing with a Schottky junction, where, even at positive voltages, the bands are still bent in the depletion region (see the band scheme in Fig. 3c). At high pumping power, the drop of the oscillation amplitude coincides with the increase of the carrier sweep-out time. With increasing carrier sweep-out time—but at constant number of carriers—the maximum magnetic field strength decreases, and thus the oscillation amplitude should decrease as observed. For low pumping power, only a very weak decrease of the amplitude can be picked up in our experiment, coinciding with a moderate increase of the sweep-out time.

We conclude that this Schottky diode works as a voltage-dependent 'magnetic field pump'. The most straightforward application of this approach to ultrafast production of magnetic fields is in the emerging experimental field of precessional dynamics. Because of the high degree of integration, our device avoids all the usual constructions required when a magnetic field that has been generated in an ultrafast fashion is brought to a magnetic element from a distant source. The field pulse is not distorted by reflections in mismatched electrical waveguides that transport current pulses from the source to a magnetic element, as the pulse is generated within the magnetic element itself. As the maximum induced magnetic field scales with the reciprocal of the spot size, reducing the size of the light spot or using hybrid junctions laterally patterned to nanometre diameters might produce sufficiently strong fields to excite precessional switching. □

Received 23 August; accepted 27 September 2001.

- Hiebert, W. K., Stankiewicz, A. & Freeman, M. R. Direct observation of magnetic relaxation in a small permalloy disk by time-resolved scanning Kerr microscopy. *Phys. Rev. Lett.* **79**, 1134–1137 (1997).
- Crawford, T. M., Silva, T. J., Teplin, C. W. & Rogers, C. T. Subnanosecond magnetization dynamics measured by the second-harmonic magneto-optic Kerr effect. *Appl. Phys. Lett.* **74**, 3386–3388 (1999).
- Hicken, R. J. & Wu, J. Observation of ferromagnetic resonance in the time domain. *J. Appl. Phys.* **85**, 4580–4582 (1999).
- Acremann, Y. *et al.* Imaging precessional motion of the magnetization vector. *Science* **290**, 492–495 (2000).
- Bauer, M., Lopusnik, R., Fassbender, J. & Hillebrands, B. Suppression of magnetic-field pulse-induced magnetization precession by pulse tailoring. *Appl. Phys. Lett.* **76**, 2758–2760 (2000).
- Choi, B. C., Belov, M., Hiebert, W. K., Ballentine, G. E. & Freeman, M. R. Ultrafast magnetization reversal dynamics investigated by time domain imaging. *Phys. Rev. Lett.* **86**, 728–731 (2001).
- Back, C. H. *et al.* Minimum field strength in precessional magnetization reversal. *Science* **285**, 864–867 (1999).
- Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
- Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
- Weber, W., Riesen, S. & Siegmund, H. C. Magnetization precession by hot spin injection. *Science* **291**, 1015–1018 (2001).
- Tsoi, M. *et al.* Excitation of a magnetic multilayer by an electric current. *Phys. Rev. Lett.* **80**, 4281–4284 (1998).
- Myers, E. B. *et al.* Current-induced switching of domains in magnetic multilayer devices. *Science* **285**, 867–870 (1999).
- Wegrowe, J.-E. *et al.* Current-induced magnetization reversal in magnetic nanowires. *Europhys. Lett.* **45**, 626–670 (1999).
- Fiederling, R. *et al.* Injection and detection of a spin-polarized current in a light-emitting diode. *Nature* **402**, 787–789 (1999).
- Ohno, Y. *et al.* Electrical spin injection in a ferromagnetic semiconductor heterostructure. *Nature* **402**,

790–792 (1999).

- Zhu, J.-G., Zheng, Y. & Prinz, G. A. Ultrahigh density vertical magnetoresistive random access memory. *J. Appl. Phys.* **87**, 6668–6673 (2000).
- Prinz, G. A., Rado, G. T. & Krebs, J. J. Magnetic properties of single-crystal (110) iron films grown on GaAs by molecular beam epitaxy. *J. Appl. Phys.* **53**, 2087–2091 (1982).
- Dumm, M. *et al.* Magnetism of ultrathin FeCo (001) films on GaAs(001). *J. Appl. Phys.* **87**, 5457–5459 (2000).
- Kittel, C. On the theory of ferromagnetic resonance absorption. *Phys. Rev.* **73**, 155–161 (1948).
- Koopmans, B., Van Kampen, M., Kohlhepp, J. T. & de Jonge, W. J. M. Ultrafast magneto-optics in nickel: magnetism or optics? *Phys. Rev. Lett.* **85**, 844–847 (2000).
- Ganping, Yu *et al.* Ultrafast time resolved photoinduced magnetization rotation in a ferromagnetic/antiferromagnetic exchange coupled system. *Phys. Rev. Lett.* **82**, 3705–3708 (1999).
- Christianen, P. C. M. *et al.* Ultrafast carrier dynamics at a metal-semiconductor interface. *J. Appl. Phys.* **80**, 6831–6838 (1996).

Acknowledgements

This work was supported by ETH Zurich, the Swiss National Foundation and the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to D.P. (e-mail: pscia@solid.phys.ethz.ch).

Imaging the atomic arrangements on the high-temperature reconstructed $\alpha\text{-Al}_2\text{O}_3(0001)$ surface

Clemens Barth & Michael Reichling

Department Chemie, Universität München, Butenandtstraße 5-13 E, 81377 München, Germany

Alumina is a technologically important oxide crystal because of its use as a catalyst and as a substrate for microelectronic applications¹. A precise knowledge of its surface atomic structure is a prerequisite for understanding and controlling the physical processes involved in many of its applications². Here we use a dynamic scanning force microscopy technique³ to image directly the atomic structure of the high-temperature phase of the $\alpha\text{-Al}_2\text{O}_3(0001)$ surface. Evidence for a surface reconstruction appears as a grid of protrusions that represent a rhombic unit cell, and we confirm that the arrangement of atoms is in the form of surface domains with hexagonal atomic order at the centre and disorder at the periphery. We show that, on exposing the surface to water and hydrogen, this surface structure is important in the formation of hydroxide clusters. These clusters appear as a regular pattern of rings that can be explained by self-organization processes involving cluster–surface and cluster–cluster interactions. Alumina has long been regarded as the definitive test for atomic-resolution force microscopy of insulators so the whole class of insulating oxides should now open for direct atomic-scale surface investigations.

The (0001) surface of $\alpha\text{-Al}_2\text{O}_3(0001)$ exists in several ordered phases that can reversibly be transformed into each other by thermal treatment and oxygen exposure⁴. The high-temperature phase has long been known to be oxygen-deficient⁵, and electron⁶ as well as X-ray diffraction⁷ revealed a surface reconstruction with a large $(\sqrt{31} \times \sqrt{31})R \pm 9^\circ$ unit cell. (Here $R \pm 9^\circ$ denotes rotation by $\pm 9^\circ$). The details and topography of this atomic structure were, however, never unambiguously revealed. Although a cubic arrangement of surface Al atoms was concluded from early low-energy electron diffraction studies⁴, results from a later, very elaborate, crystallographic study of X-ray diffraction data pointed to a hexagonal symmetry⁷ and a model was developed proposing a surface structure composed of hexagonal domains with perfect

hexagonal order in the centre and disorder at the domain boundaries^{7,8}. As alumina is a good insulator, an investigation of its surface with scanning tunnelling microscopy is impossible, so scanning force microscopy (SFM) appears to be the only technique capable of directly imaging this surface. So far, however, imaging capabilities were limited to the nanometre scale rather than atomic resolution^{9–12}.

Recently, we developed dynamic-mode SFM to yield atomic resolution on fluoride surfaces¹³ and here we apply this technique to $\alpha\text{-Al}_2\text{O}_3(0001)$ with the aim of clarifying the structural details of this surface on the atomic scale. Dynamic-mode SFM primarily measures the frequency detuning of the oscillation of a cantilever that is due to the interaction of the attached sensing tip with the surface¹⁴. In standard operation, a high gain feedback loop is installed, regulating the tip–surface distance according to the frequency detuning to obtain an image that is related to surface topography. For the measurements reported here, however, the loop gain for the topography feedback was set to a comparably low value in order also to obtain a clear contrast in the frequency detuning signal¹⁵ while at the same time gaining information about surface topography¹⁶. The $(\sqrt{31} \times \sqrt{31})R + 9^\circ$ recon-

structed $\alpha\text{-Al}_2\text{O}_3(0001)$ surface was prepared using standard procedures¹⁷ involving polishing, etching and cycle-heating the crystal to temperatures above 1,300 °C in ultrahigh vacuum.

In Fig. 1 we show the surface structure shortly after preparation. The dominant feature is a rhombic grid representing the unit cell of the $(\sqrt{31} \times \sqrt{31})R + 9^\circ$ reconstruction. The topography image (not shown here) reveals that the grid is protruding from the crystal. The displacement of atoms out of the surface plane is largest at the edges of the reconstruction rhombi, and contrast from atomic rows in $[11\bar{2}0]$ direction is best developed in their interiors. More details of the atomic structures can be inferred from the higher-magnification image shown in Fig. 2a that was taken at a reduced tip–surface distance. In this image, the grid structure is only in faint contrast; however, the atomic rows and individual surface atoms can clearly be identified. Image analysis illustrated in Fig. 2b reveals that each side of a rhomb is intersected by ten atomic rows.

This finding is perfectly compatible with the suggested hexagonal atomic structure of the surface^{7,8} and excludes the earlier proposed cubic structure that would imply 12 intersecting atomic rows⁴. A hexagonal arrangement of atoms can directly be identified in the interiors of three rhombi but is partly suppressed by a defect in the fourth rhomb. We note, in this and similar images, that such hexagonal arrangements of atoms are well developed almost exclusively in the centre of the rhombi—but these arrangements vanish at their sides and edges. This feature is also well in agreement with the proposed surface structure that predicts order in the centre of hexagonal surface domains and disorder at the domain boundaries^{7,8}.

The proposed hexagonal surface domains and the rhombic structure discovered here can be seen to be compatible when both structure models are superimposed as shown in Fig. 3. Ordered regions of rhombi and hexagons coincide while the sides and specifically the edges of the rhombi intersect regions where disorder is expected. Hence, with dynamic scanning force microscopy we could not only directly demonstrate a reconstruction on an insulator that is as large as that of the well known Si(111)7 × 7 surface, but also unambiguously establish hexagonal atomic order on the

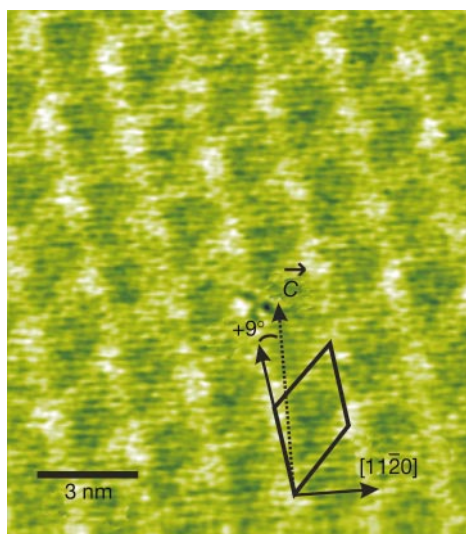


Figure 1 Structure of the $\alpha\text{-Al}_2\text{O}_3(0001)$ surface in its $(\sqrt{31} \times \sqrt{31})R + 9^\circ$ high-temperature reconstruction, as measured by dynamic scanning force microscopy. The reconstruction appears in form of a rhombic grid of protrusions with the highest elevations at the crossing points. The rhombi directly represent the unit cell of the reconstructed surface and are tilted by 9° with respect to the c -axis of the alumina crystal. Stripes in the $[11\bar{2}0]$ direction represent rows of aluminium atoms that are the predominant species on this surface. The colour-coded image represents the interaction strength (cantilever oscillation frequency detuning ranging from -44.5 Hz (dark) to -48.5 Hz (bright)) of the force microscope tip with the surface as a function of position. For this and all other measurements, the loop gain of the tip–surface distance control loop was set to a low value to yield weak topographic contrast and maintain a strong detuning signal providing the best contrast for identifying surface structure details. Atomic resolution imaging requires a well prepared tip, minimization of long-range background forces and avoidance of inhomogeneously charged surface areas. The oxidized silicon tip is prepared by repeatedly bringing it into slight contact with the surface and testing the resolution after each try. After preparing a tip yielding atomic resolution, we took great care to preserve the tip atomic structure by avoiding further contact or an excessively close approach to the surface. Atomic contrast is due to chemical interactions between a tip atom and surface ions, and can be separated from the contrast due to background forces²². To enhance chemical interaction, the tip–surface distance is manually reduced to the smallest possible values allowing imaging without tip instability. To minimize long-range electrostatic interactions between tip and surface, the bias voltage between the tip and the sample holder is varied and finally fixed at a value yielding minimum frequency detuning²².

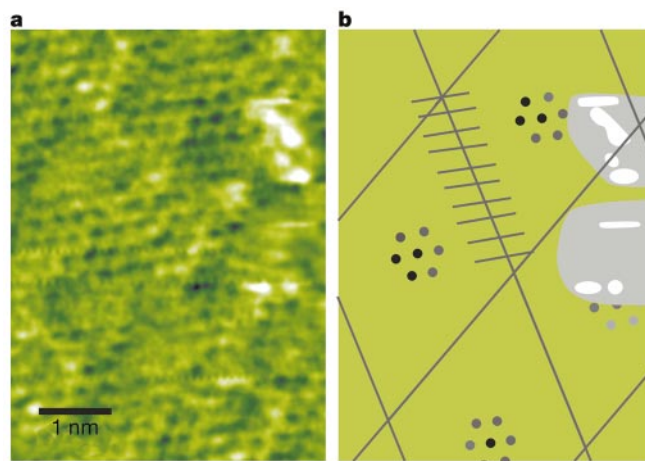


Figure 2 Section of the surface from Fig. 1 scanned with higher magnification (**a**) and schematic representation of image features shown on the same scale (**b**). The magnifying scan was performed at a reduced tip–surface distance that partly suppresses the grid structure but emphasizes the contrast due to individual surface atoms; frequency detuning ranges from -53.5 Hz (dark) to -58.5 Hz (bright). Hexagonal surface order is directly evident from the hexagons most clearly visible in the centres of the reconstruction rhombi but also from the fact that the side of each rhomb is intersected by ten atomic rows (short grey lines in **b**). This image also reveals disorder in regions far from the rhomb centres. Bright features on the right side are due to surface defects or adsorbates.

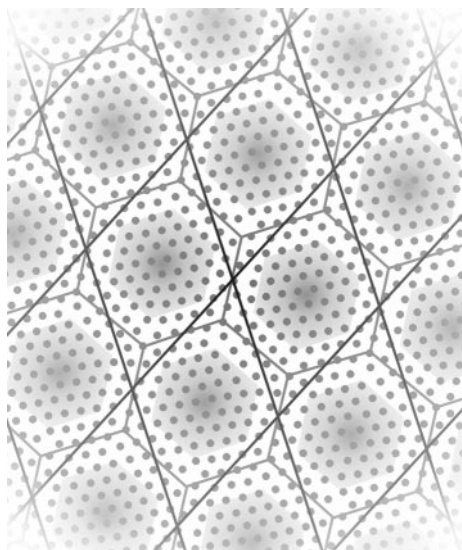


Figure 3 Schematic representation of the $(\sqrt{31} \times \sqrt{31})R + 9^\circ$ reconstruction of $\alpha\text{-Al}_2\text{O}_3(0001)$ with grey dots representing the atoms of the first aluminium layer⁸. The hexagonal pattern reflects onto the domain structure where the aluminium atoms exhibit the nearly perfect hexagonal (111) structure of metallic aluminium^{7,8}. The rhombic grid is that found by force microscopy imaging. Bright regions at the periphery of the concentric hexagons and rhombi are those of atomic disorder, while the grey shaded areas in the centre are regions of hexagonal order.

high-temperature reconstructed surface of $\alpha\text{-Al}_2\text{O}_3(0001)$ and confirm propositions about partial disorder on this surface.

Next, we demonstrate that this rhombic structure may strongly affect adsorption, clustering and crystallization at the surface. After investigating the clean, reconstructed surface, we left it exposed to residual gas that had a pressure of 4×10^{-8} Pa and was predominantly composed of water and hydrogen. Additional scanning force images were taken after dosages of 5×10^{-4} Pa s and 35×10^{-4} Pa s. Even in images (not shown here) taken at the smaller dosage, we observe a significant modification of the surface in form of defects with lateral dimensions ranging from 0.4 nm to 0.8 nm that are preferentially located in the edges of the rhombi. The number of such defects increases with exposure time and after about two days we find an unusual, regular structure of rings composed of up to eight defects, as exhibited in Fig. 4. This figure shows the rings forming a network reproducing the rhomboidal structure of the reconstructed surface. From the preferential development of defects in rhomb corners observed at lower exposures we conclude that the rings are centred at corners. This implies that the defects are preferentially stabilized in regions of disorder and maximum protrusion, which seems plausible since there we would intuitively expect lower coordination and higher chemical activity than in regions with a more regular atomic structure. This ordering into rings indicates a significant interaction between defects.

The size of the defects implies that they are not single molecules but clusters formed from species appearing when the surface is dosed with water. Their chemical composition, however, cannot be determined with the scanning force microscope and remains a subject for speculation at this stage of investigation. We find strong evidence in the literature for dissociative adsorption of water on the non-reconstructed $\alpha\text{-Al}_2\text{O}_3(0001)$ surface¹⁸ and it has been pointed out that the aluminium-terminated surface is highly reactive¹⁹. The high-temperature-reconstructed surface that we investigate here is even more aluminium-enriched, so we anticipate that upon exposure to water, the surface rapidly becomes covered with hydroxyl groups. On the other hand, it has been proposed that hydroxyl groups have significant mobility on the

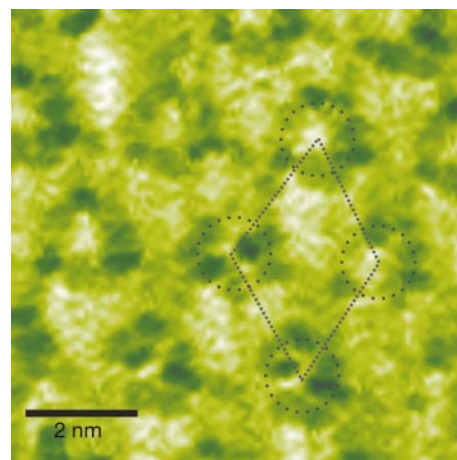


Figure 4 The reconstructed $\alpha\text{-Al}_2\text{O}_3(0001)$ surface after exposure to 35×10^{-4} Pa s of water and hydrogen originating from the residual gas of the ultrahigh vacuum. Hydroxide crystallizes into a pattern reproducing the rhomboidal structure of the reconstructed surface. A substructure in form of rings is clearly visible. Rings are centred at the points of highest elevation and largest disorder of the reconstructed surface, and are formed by up to 8 clusters. Each cluster has a size ranging between 0.4 nm and 0.8 nm and is an agglomerate of hydroxide, most probably in the form of crystalline $\text{Al}(\text{OH})_3$. In this image, a dark contrast in the frequency detuning (ranging from -45 Hz (dark) to -65 Hz (bright)) corresponds to protrusion in topography. This contrast inversion (compared to Figs 1 and 2) and the much larger detuning range are presumably due to the different chemical interactions between the tip and Al_2O_3 and between the tip and the OH clusters. It seems that the distance dependence of tip–surface interaction differs considerably between materials, resulting in a strong chemical contrast when crossing from one material to the other.

surface²⁰ and can thus condense into clusters at energetically favourable sites. As the shape of individual clusters found in our measurements is surprisingly sharp and regular, we propose that they represent small crystallites of $\text{Al}(\text{OH})_3$. In fact, there is experimental evidence for the formation of this species on alumina surfaces under high-pressure conditions²¹ and when the surface is hydrated (by rinsing with water)¹⁹. The extremely high reactivity expected in regions of atomic disorder on our surface readily explains the observed stabilization of $\text{Al}(\text{OH})_3$ clusters in these regions even under low-pressure conditions.

The ring structure, however, is determined by cluster–cluster interactions and presumably also steric constraints for their arrangement. Thus strong cluster–surface and cluster–cluster interactions create both an overall structure that corresponds to the surface reconstruction and a substructure in the form of clusters grouped in rings that has no evident relation to surface symmetry. These findings for high-temperature-reconstructed, water-exposed $\alpha\text{-Al}_2\text{O}_3(0001)$ reveal the unusual ability of this surface to structure molecular clusters formed by self-organization of the randomly positioned molecules. □

Received 18 July; accepted 13 September 2001.

1. Heinrich, V. E. & Cox, P. A. *The Surface Science of Metal Oxides* (Cambridge Univ. Press, Cambridge, 1994).
2. Campbell, C. T. Ultrathin metal films and particles on oxide surfaces: Structural, electronic and chemisorptive properties. *Surf. Sci. Rep.* **27**(1–3), 1–111 (1997).
3. de Lozanne, A. Atomic force microscopy: You may squeeze the atoms but don't mangle the surface! *Science* **291**, 2561–2562 (2001).
4. French, T. M. & Somorjai, G. A. Composition and surface structure of the (0001) face of α -alumina by low-energy electron diffraction. *J. Phys. Chem.* **74**, 2489–2495 (1970).
5. Gautier, M., Duraud, J. P., Pham Van, L. & Guittet, M. J. Modifications of $\alpha\text{-Al}_2\text{O}_3(0001)$ surfaces induced by thermal treatments or ion bombardment. *Surf. Sci.* **250**, 71–80 (1991).
6. Chang, C. C. LEED studies of the (0001) face of α -alumina. *J. Appl. Phys.* **39**, 5570–5573 (1968).
7. Renaud, G., Villette, B., Vilfan, L. & Bourret, A. Atomic structure of the $\alpha\text{-Al}_2\text{O}_3(0001)$ ($\sqrt{31} \times \sqrt{31}R \pm 9^\circ$) reconstruction. *Phys. Rev. Lett.* **73**, 1825–1828 (1994).
8. Vilfan, L., Lançon, F. & Villain, J. Rotational reconstruction of sapphire (0001). *Surf. Sci.* **392**, 62–68 (1997).

9. Beitel, G., Markert, K., Wiechers, J., Hrbek, J. & Behm, R. J. in *Adsorption on Ordered Surface of Ionic Solids and Thin Films* (eds Umbach, E. & Freund, H.-J.) 71–82 (Springer, Berlin, Heidelberg, 1993).
10. Pham Van, L., Kurnosikov, O. & Cousty, J. Evolution of steps on vicinal (0001) surfaces of α -alumina. *Surf. Sci.* **411**, 263–271 (1998).
11. Pang, C. L., Raza, H., Haycock, S. A. & Thornton, G. Growth of copper and palladium on α - Al_2O_3 (0001). *Surf. Sci.* **460**, L510–L514 (2000).
12. Sasahara, A., Utesuka, H. & Onishi, H. Noncontact-mode atomic force microscopy of α - Al_2O_3 (0001) surface. *Jpn J. Appl. Phys.* **39**, 3773–3776 (2000).
13. Reichling, M. & Barth, C. Scanning force imaging of atomic size defects on the CaF_2 (111) surface. *Phys. Rev. Lett.* **83**(4), 768–771 (1999).
14. Albrecht, T. R., Grütter, P., Horne, D. & Rugar, D. Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity. *J. Appl. Phys.* **69**, 668–673 (1991).
15. Foster, A. S., Barth, C., Shluger, A. L. & Reichling, M. Unambiguous interpretation of atomically resolved force microscopy images of an insulator. *Phys. Rev. Lett.* **86**, 2373–2376 (2001).
16. Barth, C. & Reichling, M. Resolving ions and vacancies at step edges on insulating surfaces. *Surf. Sci. Lett.* **470**, L99–L103 (2000).
17. Toofan, J. & Watson, P. R. The termination of the α - Al_2O_3 (0001) surface: a LEED crystallography determination. *Surf. Sci.* **401**, 162–172 (1998).
18. Coustet, V. & Jupille, J. Hydroxyl groups on oxide surfaces. *Il Nuovo Cimento* **19**, 1657–1664 (1997).
19. Eng, P. J. *et al.* Structure of the hydrated α - Al_2O_3 (0001) surface. *Science* **288**, 1029–1033 (2000).
20. Hass, K. C., Schneider, W. F., Curioni, A. & Andreoni, W. The chemistry of water on alumina surfaces: reaction dynamics from first principles. *Science* **282**, 265–268 (1998).
21. Liu, P., Kendelewicz, T., Brown, G. E. Jr, Nelson, E. J. & Chambers, S. A. Reaction of water vapor with α - Al_2O_3 (0001) and α - Fe_2O_3 (0001) surfaces: synchrotron X-ray photoemission studies and thermodynamic calculations. *Surf. Sci.* **417**, 53–65 (1998).
22. Guggisberg, M. *et al.* Separation of interactions by noncontact force microscopy. *Phys. Rev. B* **61**, 11151–11155 (2000).

Acknowledgements

We are grateful to A. Foster, H. Knözinger, M. Mozaffari-Afshar, Th. Risse, A. L. Shluger, M. Stoneham, I. Vilfan and K. Yasuda for discussions. The alumina sample was provided by the Fritz-Haber-Institut Berlin and technical support from this side for sample preparation is gratefully acknowledged.

Correspondence and requests for materials should be addressed to M.R.
(e-mail: reichling@cup.uni-muenchen.de).

Photoselective adaptive femtosecond quantum control in the liquid phase

T. Brixner, N. H. Damrauer, P. Niklaus & G. Gerber

Physikalisches Institut, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

Coherent light sources can be used to manipulate the outcome of light–matter interactions by exploiting interference phenomena in the time and frequency domain. A powerful tool in this emerging field of ‘quantum control’^{1–6} is the adaptive shaping of femtosecond laser pulses^{7–10}, resulting, for instance, in selective molecular excitation. The basis of this method is that the quantum system under investigation itself guides an automated search, via iteration loops, for coherent light fields best suited for achieving a control task designed by the experimenter¹¹. The method is therefore ideal for the control of complex experiments^{7,12–20}. To date, all demonstrations of this technique on molecular systems have focused on controlling the outcome of photo-induced reactions in identical molecules, and little attention has been paid to selectively controlling mixtures of different molecules. Here we report simultaneous but selective multi-photon excitation of two distinct electronically and structurally complex dye molecules in solution. Despite the failure of single parameter variations (wavelength, intensity, or linear chirp control), adaptive femtosecond pulse shaping can reveal complex laser fields to achieve chemically selective molecular excitation. Furthermore, our results prove that phase coherences of the solute molecule persist for more than 100 fs in the solvent environment.

The experimental set-up is shown schematically in Fig. 1. We investigated the use of single-parameter quantum control schemes (variation of wavelength, intensity, and linear chirp) and many-parameter ‘adaptive’ control. The latter methodology uses a femtosecond pulse shaper^{8,10,21} capable of producing complex shapes by spectral phase modulation. An automated ‘learning loop’⁸ is then iterated, in which an experimental feedback signal from the physical control subject itself guides the algorithm to find laser fields (within a 128-parameter search space) optimized for the control task designed by the experimenter. Cycling through this loop does not require any mechanistic information about the induced photo-processes, and optimum results are found under the given experimental conditions. In our experiments, laser pulses modified in single- or multi-parameter schemes irradiate methanol solutions of DCM (4-dicyanomethylene-2-methyl-6-*p*-dimethylaminostyryl-4H-pyran) and $[\text{Ru}(\text{dpp})_3](\text{PF}_6)_2$ (where dpp is 4,4’-diphenyl-2,2’-bipyridine). The linear absorption spectra (Fig. 2a) show that electronic excitation requires the absorption of at least two photons at 800 nm. Following excited-state dynamics, emissive charge-separated states are reached in both molecules^{22–24}. As these target states spontaneously decay back to the ground state after the interaction with the femtosecond laser pulse is over, their relative emission yields can be used as a measure of the excited-state populations that have been generated by the excitation pulse. The following series of experiments were designed to find out if, and how, coherent light fields could be used to excite selectively one specific molecule (but not another one) within liquid-phase solute/

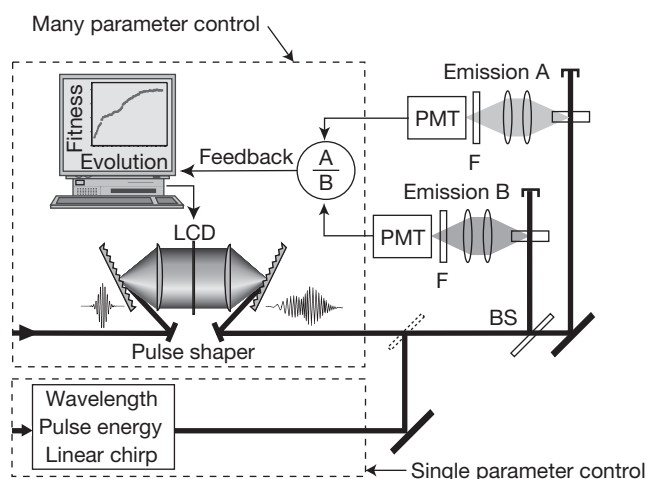


Figure 1 Experimental set-up. Our femtosecond laser system delivers 80-fs, 1-mJ pulses at a centre wavelength of 800 nm and a repetition rate of 1 kHz. Regarding single-parameter control schemes, the laser power can be varied by attenuation with neutral-density filters, and a spectral transmission window or variable linear chirp can be introduced with a femtosecond pulse shaper^{8,10}. This pulse shaper is also used in the many-parameter control scheme to generate complex electric fields. The laser spectrum is dispersed onto a liquid-crystal display (LCD). By applying suitable voltages to 128 independent LCD pixels, optical path lengths can be adjusted for the spatially separated wavelength components, leading to a modulated temporal intensity and phase profile while the spectral intensity distribution remains unchanged. The modified laser pulses are then separated by a beam splitter (BS) and passed unfocused into two flow cells (2 mm length), each containing solutions of different molecular species—shown here as A and B. The emission signals are collected at approximate right angles, spectrally filtered (F) at 650 nm for DCM (A) and at 630 nm for $[\text{Ru}(\text{dpp})_3]^{2+}$ (B), detected by photomultiplier tubes (PMT), and temporally integrated by boxcar averagers. The emission ratio of molecule A versus molecule B is used as a feedback signal in a learning loop. We use an evolutionary algorithm as described previously⁸, which (after random initialization) iteratively improves the applied pulse shape until an optimum is found. A population pool of 60 individuals per generation is used, out of which 10 survivors reproduce by cloning (giving rise to 10 children) as well as mutation (10 children) and random pixel crossover between randomly selected parental pairs (40 children).

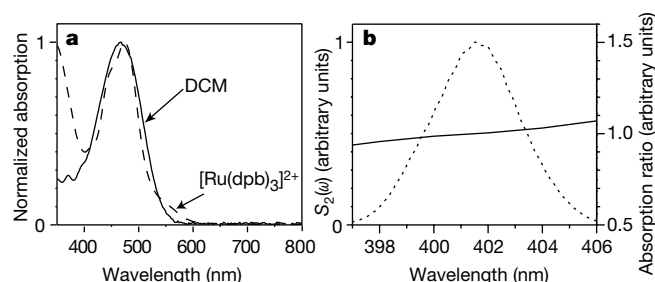


Figure 2 Spectral properties of the molecular systems. **a**, Normalized linear absorption spectra of room-temperature methanol solutions of DCM (solid line) and [Ru(dpb)₃]²⁺ (dashed line). **b**, Relative DCM/[Ru(dpb)₃]²⁺ absorption ratio (solid line), and power spectrum $S_2(\omega)$ of $E^2(t)$ (dotted line, see text for definition) of the bandwidth-limited laser pulse shown in Fig. 4a.

solvent mixtures. Specifically, the goal was to induce population transfer within DCM while simultaneously suppressing population transfer within [Ru(dpb)₃]²⁺. Hence, maximization of the DCM/[Ru(dpb)₃]²⁺ emission yield ratio was the control objective.

An obvious starting place for selective excitation of molecules using light is to choose the appropriate excitation wavelength. However, if the absorption spectra of the molecules under consideration are very similar, one-photon excitation does not select a single species. This situation is illustrated by the overlapping linear absorption spectra of the two molecules considered here (Fig. 2a). As we use broadband 800-nm femtosecond laser pulses with an electric field $E(t)$, the power spectrum of $E^2(t)$ for a bandwidth-limited laser pulse (Fig. 2b, dotted line) indicates which transition frequencies are possible in two-photon processes. The DCM/[Ru(dpb)₃]²⁺ linear absorption ratio within this spectral range (Fig. 2b, solid line) does not vary significantly, and therefore selective excitation of one molecule with respect to the other is not possible by varying the corresponding (one-photon) excitation wavelength.

We now ask if variation of laser power could be used to achieve control over the relative excitation processes. In Fig. 3a, the DCM/[Ru(dpb)₃]²⁺ emission ratio is plotted as a function of pulse energy (applying the bandwidth-limited laser pulse shown in Fig. 4a). It is seen that the excitation ratio between the two molecules remains constant. This reflects the fact that two-photon absorption is responsible for the emissive yield in both molecules, and the second-order power dependencies cancel each other when considering emission ratios. Therefore, under these conditions simple laser intensity control cannot be used to achieve selective excitation.

A third single-parameter variable frequently used in coherent control schemes is linear chirp^{25–28}. In that case, the available photon energies offered to the quantum system either increase (positive chirp, see the example in Fig. 4b) or decrease (negative chirp) linearly as a function of time within the laser pulse. Linearly chirped pulses have been used successfully in several liquid-phase control

experiments where the chromophore under study could be excited by single-photon absorption^{7,27,28}. However, under the conditions here the DCM/[Ru(dpb)₃]²⁺ emission ratio does not change as a function of linear chirp (Fig. 3b). Qualitatively, this happens because the two molecules DCM and [Ru(dpb)₃]²⁺ respond to linear chirp exactly in the same way: Introduction of chirp increases the laser pulse duration, reduces the temporal peak intensity, and therefore reduces the transition probability for the induced non-resonant two-photon excitation processes.

These results (Fig. 3a and b) are in accordance with the expectations based on previous experiments on two-photon excitation²⁹. A fourth unsuccessful single-parameter scheme—scanning the fundamental wavelength (Fig. 3c)—will be discussed below. All these results seem to indicate that the excitation processes for each molecule in the case of spectrally broad absorption bands depend only on the square of the intensity and therefore cannot be separately controlled. However, this assumption is wrong, and selective excitation can be achieved if many-parameter adaptive quantum control is employed. Using the DCM/[Ru(dpb)₃]²⁺ emission ratio as a feedback signal, the evolutionary algorithm iteratively improves the signal by approximately 50% (Fig. 3d). This shows that one molecule (DCM) can be selectively excited with respect to the other molecule ([Ru(dpb)₃]²⁺). It should be emphasized that control is possible in the presence of complex solvent/solute interactions. The resulting optimized laser field is shown in Fig. 4c. Note that the complicated structures do not reflect a random distribution of laser energy, because random pulses achieve no chemical selectivity. This can be seen from Fig. 3d, where the randomly initialized starting population of the first generation achieves a significantly lower DCM/[Ru(dpb)₃]²⁺ emission ratio. A second optimization run using a different starting configuration (constant phase initialization instead of random phases) resulted in a similar optimized emission ratio and qualitatively similar electric field. The Husimi distribution³⁰, going from high to low frequencies, first bends towards positive times, then to negative and back to positive times. This indicates that the optimized electric fields are specifically adapted to guide the temporal evolution of the system into the desired target state^{5,11}.

The scope for interpretation of the optimized control field is reduced by the explicit demonstration that certain classes of control fields do not work. This allows us to draw additional conclusions about the dynamics of the controlled quantum system. For this purpose, we first assume that the population in the molecular excited-state manifold can be calculated using second-order time-dependent perturbation theory, analogous to work^{13,29} on narrow atomic-line two-photon transitions. However, here the total excited-state population has to be calculated as a sum over the probabilities of excitation to the individual electronically excited (vibrational) sublevels. The consequences of these model assumptions and their testing against experimental evidence will be discussed below. If two-photon processes are responsible for the excitation, the power spectrum of $E^2(t)$ is relevant for the excitation. It can be expressed as²⁹

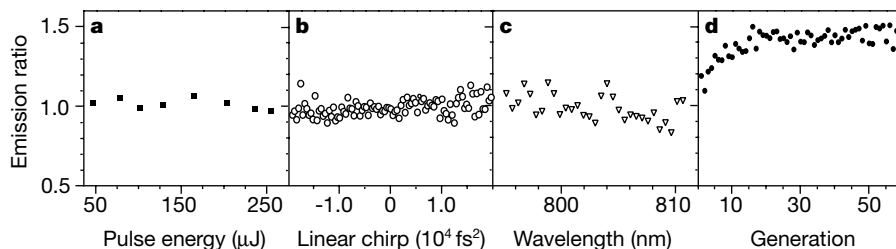


Figure 3 Control of molecular excitation. The DCM/[Ru(dpb)₃]²⁺ emission ratio is shown for: **a**, varying pulse energies of unshaped laser pulses; **b**, varying linear chirp $d^2\Phi/d\omega^2|_{\omega=\omega_0}$ where ω_0 is the centre frequency; **c**, scanning a symmetric and

rectangular spectral window of 5 nm width over the laser spectrum; and **d**, many-parameter phase shaping as a function of generation number within the evolutionary optimization algorithm.

$$S_2(\omega) = \left| \int_{-\infty}^{\infty} E(\omega/2 + \Omega) E(\omega/2 - \Omega) d\Omega \right|^2$$

where $E(\omega)$ denotes the Fourier transform of $E(t)$. In the two-photon excitation, the two photons have frequencies $\omega/2 + \Omega$ and $\omega/2 - \Omega$, respectively, such that the sum (that is, the total transition frequency) is ω . The shape of $S_2(\omega)$ can be manipulated by changing either the spectral amplitudes $|E(\omega)|$ or phases $\Phi(\omega) = -\arg E(\omega)$.

Using this simple model, it is possible to explain the experimental data discussed up to this point. Using energy variation (Fig. 3a), all frequency contributions within the laser spectrum $|E(\omega)|^2$ are attenuated by the same fraction. As $S_2(\omega)$ does not change its shape in that case, the ratio of excitation probabilities between the two different molecular species cannot be changed either. Similarly, if a gaussian-shaped laser pulse of varying linear chirp is used, it can be shown that $S_2(\omega)$ is decreased by a frequency-independent factor, and the emission ratio is again unchanged (as seen in Fig. 3b). Hence, for any two molecules where the perturbation analysis of two-photon absorption is applicable, it is not possible to achieve selective excitation by these single-parameter schemes, because the specific molecular properties cannot be exploited. The possibility for control (in the perturbation model) arises when complex spectral phases $\Phi(\omega)$ are applied, thus changing the shape of $S_2(\omega)$ owing to the interference terms in the convolution integral. Then maximization of the emission ratio (as in Fig. 3d) can be achieved by increasing the value of $S_2(\omega)$ in those spectral regions where the two-photon absorption cross-section²⁹ of DCM is relatively large and that of $[\text{Ru}(\text{dpp})_3]^{2+}$ is relatively small. Note that the spectral dependencies of the relevant two-photon cross-sections can be quite different from the behaviour of the one-photon cross-sections. This means that control with complex phase-shaped femtosecond laser pulses is possible even if the one-photon absorption spectra in the corresponding frequency range are identical and control by direct wavelength variation is not possible (as seen in Fig. 2b).

However, it was stated above that $S_2(\omega)$ can also be changed by varying the laser spectrum. If the perturbative analysis describes the molecular responses completely, then it should be possible to change the excitation ratio between the two molecules by scanning the fundamental wavelength around 800 nm—even though the wavelength scan near 400 nm employing one-photon absorption failed to change the ratio. The maximum emission ratio should be obtained where the corresponding two-photon excitation probabilities of the two molecules are the most different. This will be explored below. Note, however, that even in that case it would still be an advantage to use many-parameter optimal phase shaping instead: the magnitude of $S_2(\omega)$ can be greatly increased at any given transition frequency ω if not only the fundamental at $\omega/2$ is present but if the different excitation pathways, due to pairs of photons with energies of $\omega/2 + \Omega$ and $\omega/2 - \Omega$, are made to interfere constructively within $S_2(\omega)$ using appropriate spectral phases $\Phi(\omega)$. Thus with phase modulation the available photons are used much more

efficiently, and a higher total excited-state population can be achieved than in the case of spectral intensity reduction. However, not only the two-photon excitation probabilities are important. This is illustrated by the second optimization run, which resulted in comparable electric control fields as described above. If simply $S_2(\omega)$ were relevant, there would be no reason to expect that the same features would be found again, because many different electric field profiles could produce the same $S_2(\omega)$.

For a more rigorous analysis, we have scanned a rectangular spectral window of 5 nm in width over the laser spectrum while recording the emission signals (Fig. 3c). The wavelength region shown is determined by the available spectral bandwidth (compare Fig. 2b). The DCM/ $[\text{Ru}(\text{dpp})_3]^{2+}$ emission ratio nowhere approaches the values achieved in the many-parameter optimization. Hence, the control mechanism cannot be described by frequency-dependent two-photon cross-sections. This is a direct proof that non-trivial coherent manipulation of excited-state photophysics is responsible for control, and this in turn shows that molecular phase coherences must exist that are not destroyed by interaction with the surrounding solvent on the timescale of the optimized pulse shape (100 fs to 1 ps).

We have achieved chemically selective molecular excitation in the liquid phase by non-trivial many-parameter adaptive quantum control. No information about molecular properties was needed in the femtosecond laser pulse shape optimization procedure, and yet the evolutionary algorithm exploited specific differences in the electronic structure, vibronic structure, and dynamics which were hidden to the simpler control schemes. This enables us to gain additional insight into the molecular solvent/solute systems, which is not possible using the single-parameter spectroscopic methods alone. The presence of long-lasting phase coherences opens the door to a variety of control schemes in which coherent superpositions of molecular states are required. The adaptive technique could be used in all cases where chemically selective excitation is desired—provided that the molecular responses are not identical. Applications might be found in chemical analysis, where specific molecular properties could be made visible, or in bimolecular reaction control, where specific molecules could selectively be prepared in reactive states ('activated') for purposes of chemical synthesis. For example, consider a mixture of three distinct chemical species in solution where photoexcited species A^* or B^* undergo rapid irreversible bimolecular reactions with C to form AC or BC, respectively. In the absence of selectivity, light would produce at the very least a mixture of AC and BC. Photosensitive excitation, on the other hand, would allow for controlled production of individual bimolecular product channels (either AC or BC). Another application might be found in multi-photon fluorescence microscopy: it might be possible to obtain emission signals preferentially from a specific fluorophore, and thus spatial (microscopic) resolution could be combined with 'chemical resolution'. □

Received 29 June; accepted 10 September 2001.

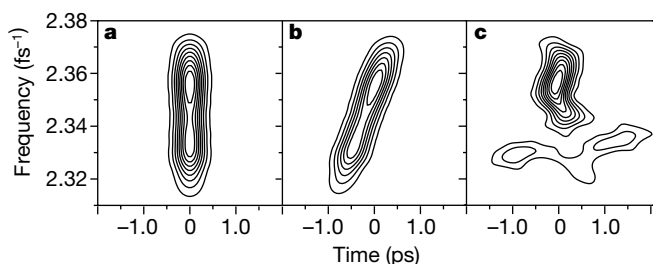


Figure 4 Experimental pulse shapes. Shown are Husimi representations³⁰ of the laser pulses determined by: **a**, the bandwidth limit; **b**, linear chirp of $2 \times 10^4 \text{ fs}^2$; and **c**, the optimized electric field from the last generation of Fig. 3d.

- Warren, W. S., Rabitz, H. & Dahleh, M. Coherent control of quantum dynamics—The dream is alive. *Science* **259**, 1581–1589 (1993).
- Rice, S. A. & Zhao, M. *Optical Control of Molecular Dynamics* (Wiley, New York, 2000).
- Rabitz, H., de Vivie-Riedle, R., Motzkus, M. & Kompa, K. Whither the future of controlling quantum phenomena? *Science* **288**, 824–828 (2000).
- Shapiro, M. & Brumer, P. Coherent control of atomic, molecular, and electronic processes. *Adv. Atom. Mol. Opt. Phys.* **42**, 287–345 (2000).
- Tannor, D. J. & Rice, S. A. Control of selectivity of chemical reaction via control of wavepacket evolution. *J. Chem. Phys.* **83**, 5013–5018 (1985).
- Brixner, T., Damrauer, N. H. & Gerber, G. Femtosecond quantum control. *Adv. Atom. Mol. Opt. Phys.* **46**, 1–54 (2001).
- Bardeen, C. J. *et al.* Feedback quantum control of molecular electronic population transfer. *Chem. Phys. Lett.* **280**, 151–158 (1997).
- Baumert, T., Brixner, T., Seyfried, V., Strehle, M. & Gerber, G. Femtosecond pulse shaping by an evolutionary algorithm with feedback. *Appl. Phys. B* **65**, 779–782 (1997).
- Yelin, D., Meshulach, D. & Silberberg, Y. Adaptive femtosecond pulse compression. *Opt. Lett.* **22**, 1793–1795 (1997).
- Brixner, T., Strehle, M. & Gerber, G. Feedback-controlled optimization of amplified femtosecond laser pulses. *Appl. Phys. B* **68**, 281–284 (1999).

11. Judson, R. S. & Rabitz, H. Teaching lasers to control molecules. *Phys. Rev. Lett.* **68**, 1500–1503 (1992).
12. Assion, A. *et al.* Control of chemical reactions by feedback-optimized phase-shaped femtosecond laser pulses. *Science* **282**, 919–922 (1998).
13. Meshulach, D. & Silberberg, Y. Coherent quantum control of two-photon transitions by a femtosecond laser pulse. *Nature* **396**, 239–242 (1998).
14. Weinacht, T. C., White, J. & Bucksbaum, P. H. Toward strong field mode-selective chemistry. *J. Phys. Chem. A* **103**, 10166–10168 (1999).
15. Weinacht, T. C., Ahn, J. & Bucksbaum, P. H. Controlling the shape of a quantum wavefunction. *Nature* **397**, 233–235 (1999).
16. Hornung, T., Meier, R. & Motzkus, M. Optimal control of molecular states in a learning loop with a parameterization in frequency and time domain. *Chem. Phys. Lett.* **326**, 445–453 (2000).
17. Kunde, J. *et al.* Adaptive feedback control of ultrafast semiconductor nonlinearities. *Appl. Phys. Lett.* **77**, 924–926 (2000).
18. Bartels, R. *et al.* Shaped-pulse optimization of coherent emission of high-harmonic soft X-rays. *Nature* **406**, 164–166 (2000).
19. Levis, R. J., Menkir, G. M. & Rabitz, H. Selective bond dissociation and rearrangement with optimally tailored, strong-field laser pulses. *Science* **292**, 709–713 (2001).
20. Brixner, T., Kiefer, B. & Gerber, G. Problem complexity in femtosecond quantum control. *Chem. Phys.* **267**, 241–246 (2001).
21. Weiner, A. M. Femtosecond pulse shaping using spatial light modulators. *Rev. Sci. Instrum.* **71**, 1929–1960 (2000).
22. Damrauer, N. H., Boussie, T. R., Devenney, M. & McCusker, J. K. Effects of intraligand electron delocalization, steric tuning, and excited-state vibronic coupling on the photophysics of aryl-substituted bipyridyl complexes of Ru(II). *J. Am. Chem. Soc.* **119**, 8253–8268 (1997).
23. Damrauer, N. H. & McCusker, J. K. Ultrafast dynamics in the metal-to-ligand charge transfer excited-state evolution of [Ru(4,4'-diphenyl-2,2'-bipyridine)]²⁺. *J. Phys. Chem. A* **103**, 8440–8446 (1999).
24. Meyer, M., Mialocq, J. C. & Perly, B. Photoinduced intramolecular charge transfer and *trans-cis* isomerization of the DCM styrene dye. Picosecond and nanosecond laser spectroscopy, high-performance liquid chromatography, and nuclear magnetic resonance studies. *J. Phys. Chem.* **94**, 98–104 (1990).
25. Assion, A., Baumert, T., Helbing, J., Seyfried, V. & Gerber, G. Coherent control by a single phase shaped femtosecond laser pulse. *Chem. Phys. Lett.* **259**, 488–494 (1996).
26. Broers, B., van Linden van den Heuvell, H. B. & Noordam, L. D. Efficient population transfer in a three-level ladder system by frequency-swept ultrashort laser pulses. *Phys. Rev. Lett.* **69**, 2062–2065 (1992).
27. Bardeen, C. J., Wang, Q. & Shank, C. V. Selective excitation of vibrational wave-packet motion using chirped pulses. *Phys. Rev. Lett.* **75**, 3410–3413 (1995).
28. Cerullo, G., Bardeen, C. J., Wang, Q. & Shank, C. V. High-power femtosecond chirped pulse excitation of molecules in solution. *Chem. Phys. Lett.* **262**, 362–368 (1996).
29. Meshulach, D. & Silberberg, Y. Coherent quantum control of multiphoton transitions by shaped ultrashort optical pulses. *Phys. Rev. A* **60**, 1287–1292 (1999).
30. Lalović, D., Davidović, D. M. & Bijedić, N. Quantum mechanics in terms of non-negative smoothed Wigner functions. *Phys. Rev. A* **46**, 1206–1212 (1992).

Acknowledgements

This work was supported by the European Coherent Control Network (COCOMO) and the Fonds der Chemischen Industrie. N.D. thanks the Alexander von Humboldt-Stiftung for a postdoctoral fellowship.

Correspondence and requests for materials should be addressed to G.G. (e-mail: gerber@physik.uni-wuerzburg.de).

Thinning of the ice sheet in northwest Greenland over the past forty years

W. S. B. Paterson*[†] & Niels Reeh*[‡]

[†] Paterson Geophysics Inc., Box 303, Heriot Bay, British Columbia VOP 1H0, Canada

[‡] Ørsted-DTU, Electromagnetic Systems, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

* The authors contributed equally to this work

Thermal expansion of the oceans, as well as melting of glaciers, ice sheets and ice caps have been the main contributors to global sea level rise over the past century. The greatest uncertainty in predicting future sea level changes lies with our estimates of the mass balance of the ice sheets in Greenland and Antarctica¹. Satellite measurements have been used to determine changes in these ice sheets on short timescales, demonstrating that surface-

elevation changes on timescales of decades or less result mainly from variations in snow accumulation². Here we present direct measurements of the changes in surface elevation between 1954 and 1995 on a traverse across the north Greenland ice sheet. Measurements over a time interval of this length should reflect changes in ice flow—the important quantity for predicting changes in sea level—relatively unperturbed by short-term fluctuations in snow accumulation. We find only small changes in the eastern part of the transect, except for some thickening of the north ice stream. On the west side, however, the thinning rates of the ice sheet are significantly higher and thinning extends to higher elevations than had been anticipated from previous studies³.

Because the Greenland ice sheet is still responding to climatic changes that occurred thousands of years ago⁴, a complicated pattern of thickening and thinning is expected. The earliest data of sufficient precision appear to be those of the British North Greenland Expedition (BNGE)⁵; we compare these with the latest digital elevation model⁶. The BNGE used trigonometric levelling to measure elevation at about 300 stations on a 1,200-km traverse in 1953–54 (Fig. 1)⁷. The survey started at Krebs Bjerg in Dronning Louise Land and ended at sea level on the west coast. The route crossed the north ice stream⁸. The work was spread over two summers; a tripod was left standing at the midpoint, station B73, throughout the winter.

The closure error, found by calculating the elevation of B73 from each coast, was 13.8 m, elevations calculated from the east being the higher. It was subsequently reduced to 11.8 m when the elevation of Krebs Bjerg was redetermined. It was distributed as follows. East of B73, each height difference between adjacent stations was reduced by 5.9 m times the distance between them, expressed as a fraction of the total distance between Krebs Bjerg and B73. Height differences west of B73 were increased in a similar way. To set error limits, for stations east of B73 the elevation calculated from the east coast was taken as the upper limit and the lower limit was taken as that value minus twice the adjustment in elevation. Similarly, for stations west of B73, the elevation calculated from the west coast set a lower limit and twice the adjustment was added to obtain the upper limit. A second scheme, with the adjustment proportional to the elevation difference, rather than the distance, between adjacent stations was also used. Elevations calculated by the two schemes differed by up to 3 m. Their mean was taken as the true elevation; the maximum error is the root mean square of the adjustments calculated for the separate schemes. The latitude and longitude of each station were found

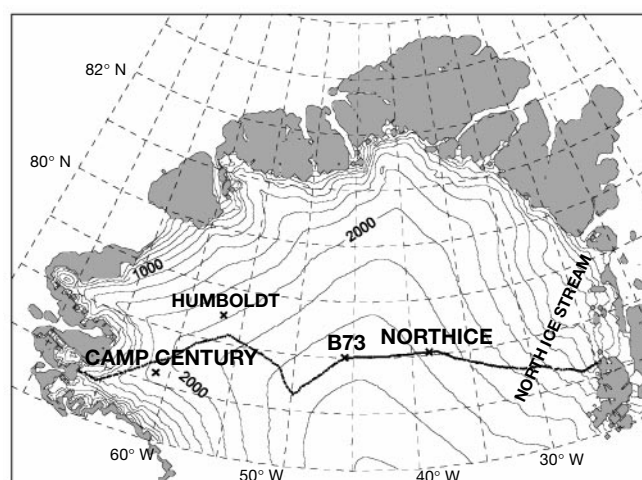


Figure 1 Traverse route. Northice was the Expedition's ice-sheet station. Long-term snow accumulation data are available for Humboldt Station and Camp Century.

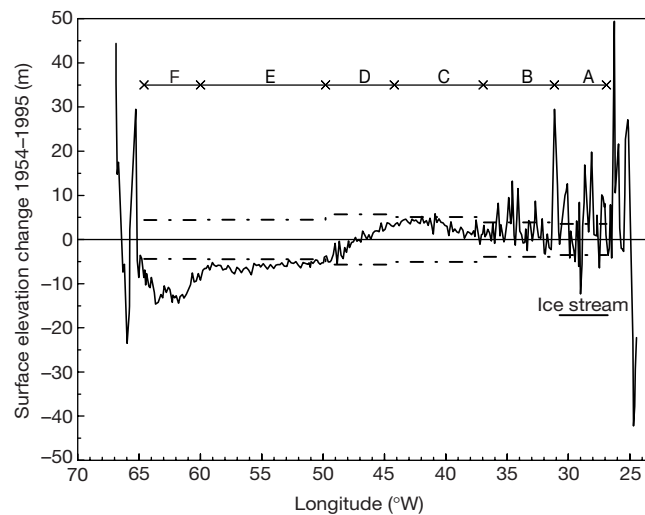


Figure 2 Measured changes in surface elevation between 1954 and 1995. Dashed-dotted lines indicate maximum errors. Letters denote longitude bands used in the data analysis.

from the measured distances and azimuths and checked by Sun sights at some stations. The precision is 0.1 minute of latitude and 0.5 minute of longitude, or about 200 m in each direction.

The digital elevation model (DEM) of the ice sheet, which has a grid spacing of 1 km, is based, in north Greenland, on radar altimetry from the ERS-1 satellite collected in 1994–95. The data were corrected for a slope-dependent bias that arises because the radar oversamples the crests of surface undulations relative to the troughs. The corrected radar data were compared with measurements of surface elevation made by laser altimeters in an aircraft. Because these have a precision of about 10 cm, the standard deviation of the difference between DEM and laser elevations, which is a function of surface slope, measures the precision of the DEM⁶. Although the laser elevations are more precise than the DEM, the data points are more scattered and some are several kilometres from the nearest BNGE station. We have therefore used the DEM.

DEM elevations are relative to the World Geodetic Systems 1984 (WGS 84) ellipsoid whereas BNGE elevations are relative to sea level. The conversion, made using the latest geoid model (Ohio State University 1991A; OSU 91A) improved with local gravity, introduces a maximum error of ± 1 m.

Figure 2 shows the measured changes in surface elevation. They can be interpreted as ice-thickness changes because bedrock beneath the ice is unlikely to be rising more than a few millimetres per year. To test the significance of the changes, we grouped the stations into longitude bands and calculated average values for each. We excluded the marginal regions, where the radar was probably measuring the elevations of nunataks off the traverse route. (A nunatak is an isolated peak of rock projecting above the surface of land ice.) The maximum error has four components: uncertainties in the BNGE elevations and in the DEM, both calculated as described above, the geoid–ellipsoid conversion, and the scatter of elevation changes within each band. The standard deviation of the DEM elevations and the standard error of the mean elevation change in each band were multiplied by two to make them comparable with the other errors, which are maximum errors. The components were combined by taking the square root of the sum of their squares.

We conclude that, between 1954 and 1995: (1) The ice stream (band A in Fig. 2) has thickened at a rate of 9.7 ± 8.4 cm yr⁻¹; (2) ice thickness has not changed significantly in bands B–D; (3) the

average thinning rate is 16.5 ± 11.0 cm yr⁻¹ in band E and 31.0 ± 10.7 cm yr⁻¹ in band F. These bands span elevations from 2,500 to 1,500 m. All limits quoted are maximum, not standard, errors.

We believe that the 41-year interval is long enough to ensure that we are measuring the dynamic response of the ice sheet rather than fluctuations in snow accumulation. Furthermore, annual accumulation data from a station on Humboldt Glacier⁹ and from near Camp Century (Fig. 1; E. Mosley-Thompson, personal communication) show no trend over this period.

The only other direct measurements of elevation changes in this area come from a study covering the whole ice sheet for the period 1994–99 (ref. 10). It shows slight thickening in bands E and F where we measured significant thinning. The thickening probably resulted from variations in snow accumulation.

An indirect method avoids the difficulties in interpreting short-term thickness changes³: the total snow accumulation over a drainage basin is compared with the ice flux out of it. This certainly measures the dynamic response of the ice sheet, but only a rough average over a large area¹¹. The Greenland study was restricted to elevations above 2,000 m by measuring fluxes at that contour. In north Greenland east of longitude 40°W an average thickening of 2 cm yr⁻¹ was obtained. This is consistent with our results except in the ice stream, a feature this method cannot resolve. Thinning was observed to the west but, even if the upper limits are used, at rates substantially less than those we measured. □

Received 6 August; accepted 18 September 2001.

- Houghton, J. T. *et al.* *Climate Change 2001: The Scientific Basis* 664–673 (Cambridge Univ. Press, Cambridge, 2001).
- McConnell, J. R. *et al.* Changes in Greenland ice sheet elevation attributed primarily to snow accumulation variability. *Nature* **406**, 877–879 (2000).
- Thomas, R. *et al.* Mass balance of the Greenland ice sheet at high elevations. *Science* **289**, 426–428 (2000).
- Huybrechts, P. & Le Meur, E. Predicted present-day evolution patterns of ice thickness and bedrock elevation over Greenland and Antarctica. *Polar Res.* **18**(2), 299–306 (1999).
- Paterson, W. S. B. Altitudes on the inland ice of north Greenland. *Medd. Grönl.* **137**(1), 1–12 (1955).
- Bamber, J. L., Ekholm, S. & Krabill, W. B. A new high-resolution digital elevation model of Greenland fully validated with airborne laser altimeter data. *J. Geophys. Res.* **106**(B4), 6733–6745 (2001).
- Paterson, W. S. B. & Slesser, C. G. M. Trigonometric levelling across the inland ice in north Greenland. *Empire Surv. Rev.* **13**, 252–261 (1956).
- Fahnestock, M., Bindshadler, R., Kwok, R. & Jezek, K. Greenland ice sheet surface properties and ice dynamics from ERS-1 SAR imagery. *Science* **262**, 1530–1534 (1993).
- Anklin, M., Bales, R. C., Mosley-Thompson, E. & Steffen, K. Annual accumulation at two sites in northwest Greenland during recent centuries. *J. Geophys. Res.* **103**(D22), 28775–28783 (1998).

10. Krabill, W. *et al.* Greenland ice sheet: high-elevation balance and peripheral thinning. *Science* **289**, 428–430 (2000).

11. Paterson, W. S. B. *The Physics of Glaciers* 3rd edn 45 (Elsevier Science, Oxford, 1994).

Acknowledgements

We thank J. Bamber, S. Ekholm, W. Krabill and E. Mosley-Thompson for providing information and unpublished data.

Correspondence and requests for materials should be addressed to N.R. (e-mail: nr@oersted.dtu.dk).

Late Cretaceous relatives of rabbits, rodents, and other extant eutherian mammals

J. David Archibald*, Alexander O. Averianov† & Eric G. Ekdale*

* Department of Biology, San Diego State University, San Diego, California 92182-4614, USA

† Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, Saint Petersburg 199034, Russia

Extant eutherian mammals and their most recent common ancestor constitute the crown group Placentalia. This taxon, plus all extinct taxa that share a more recent common ancestor with placentals than they do with Metatheria (including marsupials), constitute Eutheria¹. The oldest well documented eutherian-dominated fauna in the world is Dzharakuduk, Uzbekistan². Among eutherians that it yields is *Kulbeckia*, an 85–90-Myr-old member of Zalambdalestidae (a family of Late Cretaceous Asian eutherians)³. This extends Zalambdalestidae back by some 10 million years from sites in the Gobi Desert, Mongolia⁴. A phylogenetic analysis of well described Late Cretaceous eutherians strongly supports Zalambdalestidae, less strongly supports 'Zhelestidae' (a Late Cretaceous clade related to Tertiary ungulates), but does not support Asioryctitheria (a group of Late Cretaceous Asian eutherians). A second analysis incorporating placentals from clades that include rodents (*Tribosphenomys*), lagomorphs (*Mimotona*) and archaic ungulates (*Protungulatum* and *Oxyprimus*) strongly supports Zalambdalestidae in a clade with Glires (rabbits, rodents and extinct relatives) and less strongly 'Zhelestidae' within a clade that includes archaic ungulates ('condylarths'). This argues that some Late Cretaceous eutherians belong within the crown group Placentalia. The ages of these taxa are in line with molecularly based estimates of 64–104 Myr ago (median 84 Myr ago) for the superordinal diversification of some placentals⁵, but provide no support for a Late Cretaceous diversification of extant placental orders.

Timing of the origin of major clades of extant placental mammals remains controversial. Some molecular studies place the origin of such clades (orders) far back into the Cretaceous⁶. Most palaeobiologists argue that such clades originated in the Palaeocene or later^{7,8}. Others argue that more inclusive extant placental clades (superordinal groupings) may not be discernible even in the Late Cretaceous⁹. Some proposed exceptions are Late Cretaceous representatives of Ungulatomorpha ('Zhelestidae'¹⁰), lipotyphlans (*Paranactoides* and *Batodon*¹¹) and Glires (Zalambdalestidae¹¹). The timing of the origin and relationships of Glires are among the most controversial subjects in studies of mammal evolution.

Glires includes extinct early Tertiary taxa, as well as extant Rodentia (rodents) and Lagomorpha (rabbits and pikas). The monophyly of Glires has been repeatedly upheld⁵ but the timing of the origin of Glires is contested; on the basis of molecular studies

one of its members, Rodentia, has been argued to appear over 100 Myr ago^{6,12}. This is almost as old as the oldest eutherian fossils^{13–15}. The relationships of Glires to other eutherian taxa are also disputed; it has been argued that Glires may share a common ancestry with the Late Cretaceous Zalambdalestidae¹¹.

We present new data for the oldest known zalambdalestid, *Kulbeckia kulbecke*, from the 85–90-Myr-old Dzharakuduk fauna, Bissekty Formation, Uzbekistan². Until recently, zalambdalestids were known almost exclusively from 75-Myr-old fossils from the Gobi Desert (J. R. Wible, M. J. Novacek & G. W. Rougier, manuscript in preparation; and ref. 16). *Kulbeckia* was named in 1993, on the basis of a few teeth¹⁷. Possible affinities with zalambdalestids were not recognized until 1997 (ref. 3). During field seasons at Dzharakuduk (1997–2000), over 40 additional specimens were recovered including teeth, dentaries, petrosals, a partial skull, and postcrania (Figs 1, 2). Similarities of zalambdalestids and Glires warrant further consideration.

The hallmark of Glires is its enlarged, medially placed pair of incisors, which are procumbent, have enamel restricted to the more ventro- or dorsolabial margin of the tooth, are open-rooted, ever-growing, and have Hunter–Schreger bands¹⁸. Other mammals possess incisors that have some of these traits, but only Glires has all six character states. In the lower incisors, the zalambdalestid *Kulbeckia* possesses the first four of these states, although the enamel encompasses a greater circumference of the crown, which is a slightly more ancestral condition (Fig. 2c). We cannot demonstrate that the open-rooted lower incisor was ever-growing and scanning electron micrographs (SEMs) of the surface of the lower incisor did not reveal Hunter–Schreger bands. Although suggestive, these character states alone are not strong evidence for a phylogenetic relationship, because other eutherians possess some of these traits. When a suite of other comparisons is added, the zalambdalestid–Glires link becomes more plausible.

The two larger, anterior upper incisors and much smaller third upper incisor known for *Kulbeckia* are preserved only as roots; thus, the disposition of enamel is unknown. One (or two) smaller, more medial incisor(s) may have been present (Fig. 2a, b). Rodents possess one upper incisor on each side while lagomorphs possess two, and thus the condition in *Kulbeckia* could have been ancestral to both of these orders. *Kulbeckia* and other zalambdalestids (*Zalambdalestes* and *Barunlestes*) warrant further comparison with Glires, especially with more basal members. All have narrow, elongate snouts (Figs 1a and 2b). *Zalambdalestes* and *Barunlestes* fall between the ancestral condition in which most of the elongation is within the maxilla and the derived condition in Glires in which half or more of the elongation is in the premaxilla. Although it shows some premaxillary elongation, *Kulbeckia* is close to the ancestral condition. In the morphocline *Kulbeckia* to *Zalambdalestes* to *Barunlestes* to Glires, we see a reduction of incisors from 3, 4, or 5/4 (5 in the upper jaw, 4 in the lower jaw) to 2/3 or 3/3 to 2/3 to 1/1 or 2/1. Homologies of the four lower incisors are open to interpretation. *Kulbeckia* may retain the ancestral therian condition of four incisor sites seen in some early eutherians. In these taxa, however, the incisors are of nearly equal size, while in *Kulbeckia* the most medial incisor is six times longer and perhaps as much wider than the three following teeth. In Glires embryological evidence suggests that the medial, enlarged lower tooth is the second, lower, deciduous incisor¹⁹. Unfortunately, embryological evidence does not help in determining homologies in available fossils, and thus without ontogenetic information we use the overall anatomy of the dentitions to support the view that the medial pair of lower incisors in zalambdalestids and Glires are homologous.

In the same taxa, canines are never large, ranging from two roots above (in the upper jaw) and below (in the lower jaw) to one above and below to no canines. In *Kulbeckia* there are 4/4 premolars while later zalambdalestids have 3/4 or 4/4, and among some Glires, 3/3 are retained. Upper molars of both *Kulbeckia* and early Glires such as

Tribosphenomys and *Mimotona* have transversely expanded crowns, a metacingulum where the postmetaconular crista continues onto the metastylar lobe, and a narrower metastylar lobe on the second upper molar^{20,31} (Fig. 1b, c). In illustrations, molar conules in *Zalambdalestes* and *Barunlestes* are shown as labial nearer the paracone and metacone and at least in the former taxon are characterized as 'incipient' and 'very indistinct'¹⁶. This is unlike the more lingual condition in *Kulbeckia* and early Glires. All published specimens of *Zalambdalestes* and *Barunlestes* are worn and thus the conular positions are subject to interpretation.

The above suite of characters in *Kulbeckia*, other zalambdalestids, and Glires is suggestive of a possible phylogenetic relationship. To test this we performed phylogenetic analyses using MacClade²¹ and PAUP²², beginning with better represented Late Cretaceous eutherians. We used 70 characters including those from previous studies (J. R. Wible, M. J. Novacek & G. W. Rougier, manuscript in preparation and refs 1, 10, 23, 24) based upon upper and lower dentitions, anterior skull, dentary and petrosal (see the Supplementary Information). Only multistate characters forming a morphocline were coded as ordered: all characters were unweighted; and the branch and bound search option was used. Results for this analysis indicate a well supported (bootstrap value over 70) Zalambdalestidae. Although Zhelestidae is also supported, it is not as well supported and may include *Pavanyctoides* and *Gallolestes* (Fig. 3a). Asioryctitheria (*Asioryctes* and *Kennalestes*) is not supported, but neither is there evidence against it. The third published asioryctitheria, *Ukhaatherium*, was not included, as it was not separately coded in earlier analyses¹.

In the next analysis we added early Tertiary placental taxa belonging to clades that are argued to be linked to Late Cretaceous eutherians^{10,11} (Fig. 3b). These taxa were chosen because they are widely accepted basal members of superordinal clades within the

crown taxon Placentalia, yet they retain some ancestral eutherian character states. These included two taxa from Glires, *Tribosphenomys* and *Mimotona*, and two archaic ungulates, *Protungulatum* and *Oxyprimus*. *Tribosphenomys* is a very basal taxon within Simplicidentata, which also includes Rodentia²⁰. *Mimotona* is a well known member of Mimotonidae, which along with Lagomorpha²⁵ comprise Duplicidentata. *Protungulatum* is a basal taxon within Ungulata or within Arctocyonidae. The latter taxon has been argued to be basal to other archaic ungulates as well as Artiodactyla and Cetacea²⁶. *Oxyprimus* is a basal ungulate regarded as either an arctocyonid or hyposodontid²⁶.

Results for this analysis yield a well supported (bootstrap values over 70) 'gliriform' clade including zalambdalestids as well as a Glires clade composed of *Tribosphenomys* and *Mimotona* (Fig. 3b). Formal taxonomic revisions are beyond the scope of this work, but we note that if the relationship among zalambdalestids, *Tribosphenomys* and *Mimotona* are supported by further analyses, Zalambdalestidae is rendered paraphyletic. The same concern has been expressed for 'Zhelestidae' and hence the placement of this taxon within quotes²³. Asian 'Zhelestidae' remain monophyletic and along with other 'zhelestids' and archaic ungulates form a polytomy with *Paranyctoides* and *Gallolestes* (Fig. 3b). *Paranyctoides* and *Gallolestes* are not referred to 'Zhelestidae', but they may belong with this clade (dashed lines in Fig. 3b). Among 'zhelestids', *Alostera* groups with archaic ungulates. *Alostera* is based on isolated teeth, but does possess features of archaic ungulates not seen in other 'zhelestids', raising the possibility that it is either an archaic ungulate or the nearest sister taxon.

We follow recent taxonomic usage that defines Placentalia as the clade formed by the most recent common ancestor of living eutherians, plus all its descendants¹. Eutheria is the clade that includes Placentalia (crown group) and all taxa that share a more

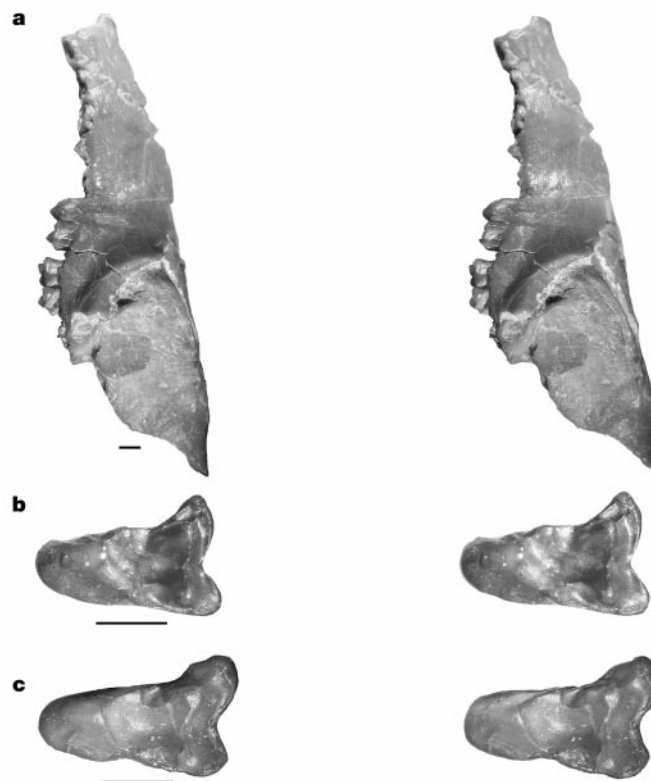


Figure 1 Stereo pairs of *Kulbeckia kulbecke*, Dzharakuduk fauna, Bissekty Formation. **a**, Left lateral view of anterior skull (URBAC 99-53); **b**, occlusal view left M1, type (CCMGE 52/12455); **c**, occlusal view left M2 (ZIN C.82565). CCMGE, Chernyshev's Central Museum of Geological Exploration, Saint Petersburg; URBAC, Uzbekistanian/Russian/

British/American/Canadian Joint Paleontological Expedition specimens in the Institute of Zoology, Tashkent, and the Royal Ontario Museum, Toronto; ZIN C., Systematic Collections, Zoological Institute, Russian Academy of Sciences, Saint Petersburg. Scale bars equal 1 mm.

recent common ancestor with it (stem eutherians) than they do with Metatheria (including the crown group Marsupialia). Metatheria is the sister taxon of Eutheria. As described above, our analysis argues that some Late Cretaceous eutherians belong within the crown group Placentalia. This view has been challenged by studies that suggest Placentalia possesses apomorphies not seen in Late Cretaceous eutherians²⁷. These derived characters are: upper incisors four or fewer, nasals not posteriorly expanded, lacrimal without well developed facial process, jugal not posteriorly extensive in the zygoma, and epipubics absent. These five characters remain poorly known for most eutherians. For example, we observed a well developed facial process of the lacrimal among both extinct and extant Artiodactyla and Perissodactyla, and extinct archaic ungulates, all of which are placentals and yet retain the argued ancestral eutherian condition. Although less common, the nasals are posteriorly expanded in various clades in the above taxa,

in some cases contacting the lacrimals. Also, the reduction or loss of epipubics^{28,29} and the reduction in number of upper incisors to four or fewer occurred in at least three different clades of metatherians³⁰. Such reductions and losses also very probably occurred many times in eutherians and thus we question their reliability for uniting all Placentalia. On the basis of our phylogenetic analyses we feel that current evidence supports a crown group, Placentalia, that includes some Late Cretaceous eutherians. More specifically, these results support a superordinal clade including zalambdalestids and Glires. A zhelestid and ungulate superordinal clade also continues to be supported.

Finally, our analysis provides corroborating evidence concerning the timing of superordinal diversification. Both zalambdalestids and zhelestids occur in the 85–90-Myr-old faunas at Dzharakuduk².

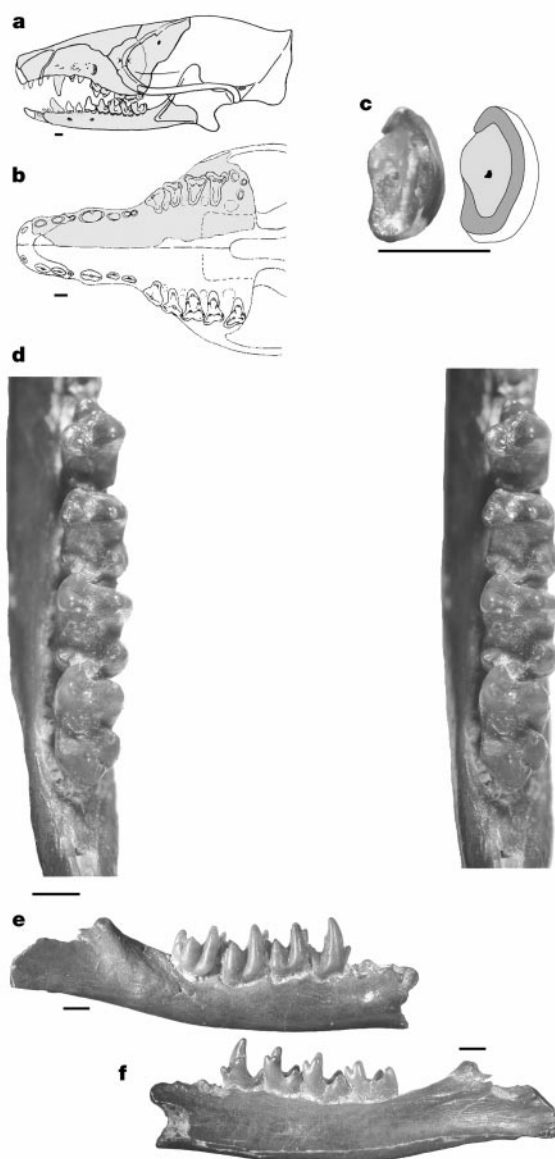


Figure 2 *Kulbeckia kulbecke*, Dzharakuduk fauna, Bissekty Formation. **a, b**, Left lateral and palatal reconstructions, with parts known shown in grey; **c**, broken tip of lower left incisor showing restriction of enamel (medium grey) to dorsal, labial and ventral surfaces (URBAC 98-3); **d**, stereo occlusal view, p5, m1-3 (URBAC 98-2); **e, f**, labial and lingual views of the same specimen. For institutional abbreviations see Fig. 1. Scale bars equal 1 mm.

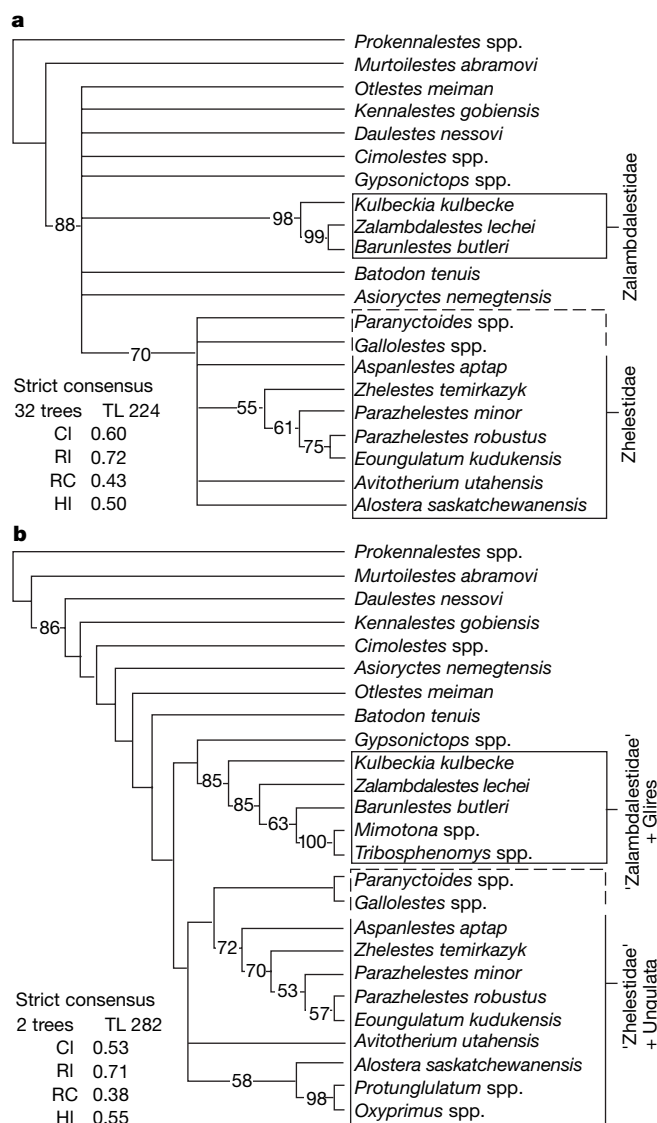


Figure 3 Phylogenetic analyses discussed in the text. Numbers refer to bootstrap values over 50. Matrix of taxa, characters, and states are defined in Supplementary Information. **a**, Cladogram of all better described Late Cretaceous eutherians with Early Cretaceous *Prokennalestes* and *Murtoilestes* as outgroup taxa, showing monophyly of Zalambdalestidae and Zhelestidae. **b**, Cladogram of same taxa but including two Tertiary members of Glires, *Mimotona* and *Tribosphenomys*, and two Tertiary members of Ungulata, *Protungulatum* and *Oxyprimus*, showing a gliiriform clade including 'Zalambdalestidae' and Glires and an ungulatormorph clade including 'Zhelestidae' and Ungulata. TL, treelength; CI, consistency index; RI, retention index; RC, rescaled consistency index; HI, homoplasy index.

The presence of zalambdalestids argues that the superordinal clade including Glires had separated from other superordinal placental clades by this time. This is also applicable for zhelestids, thus suggesting that some ungulate clades had separated from other superordinal placental clades by this time. The dates of these fossil taxa are concordant with molecularly based estimates of 64–104 Myr ago (median 84 Myr ago) for the superordinal diversification of placentals⁵. No members of extant placental orders, however, are known from the Late Cretaceous (with the possible exception of some insectivores). Subsequent diversification of living placental orders within these Late Cretaceous placental superordinal groups did not begin until about 65 Myr ago, after dinosaur extinction^{7,8}. □

Received 1 August; accepted 26 September 2001.

1. Rougier, G. W., Wiblé, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).
2. Archibald, J. D. *et al.* Précis of the paleontology, biostratigraphy, and sedimentology at Dzharakuduk (Turonian?–Santonian), Kyzylkum Desert, Uzbekistan. *New Mex. Mus. Nat. Hist. Sci. Bull.* **14**, 21–28 (1998).
3. Nessov, L. A. *Cretaceous Nonmarine Vertebrates of Northern Eurasia* 168–169 (Univ. St. Petersburg Institute of Earths Crust, St Petersburg, 1997).
4. Kielan-Jaworowska, Z., Novacek, M. J., Trofimov, B. A. & Dashzeveg, D. in *The Age of Dinosaurs in Russia and Mongolia* (eds Benton, M. J. *et al.*) 573–626 (Cambridge Univ. Press, Cambridge, 2000).
5. Murphy, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614–618 (2001).
6. Kumar, S. & Hedges, B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
7. Foote, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. Jr Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283**, 1310–1314 (1999).
8. Archibald, J. D. & Deutschman, D. H. Quantitative analysis of the timing of the origin and diversification of extant placental orders. *J. Mamm. Evol.* **8**, 107–124 (2001).
9. Novacek, M. J., Gao, K., Norell, M. A. & Rougier, G. Ghost lineages, phylogeny, and ranges of selected vertebrate lineages across the K/T boundary. *J. Vert. Paleontol. Abstr.* **18** (suppl. 3), 67A (1998).
10. Archibald, J. D. Fossil evidence for a Late Cretaceous origin of “hoofed” mammals. *Science* **272**, 1150–1153 (1996).
11. McKenna, M. C. & Bell, S. K. *Classification of Mammals above the Species Level* 104–105, **284** (Columbia Univ. Press, New York, 1997).
12. Cooper, A. & Forney, R. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* **13**, 151–156 (1998).
13. Kielan-Jaworowska, Z. & Dashzeveg, D. Eutherian mammals from the Early Cretaceous of Mongolia. *Zool. Scripta* **18**, 347–355 (1989).
14. Cifelli, R. L. Tribosphenic mammals from the North American Early Cretaceous. *Nature* **401**, 363–366 (1999).
15. Averianov, A. O. & Skutschas, P. A new genus of eutherian mammal from the Early Cretaceous of Tranbaikalia, Russia. *Acta Palaeontol. Pol.* **46**, 431–436 (2001).
16. Kielan-Jaworowska, K. Preliminary data on the Upper Cretaceous eutherian mammals from Bayn Dzak, Gobi Desert. *Palaeontol. Pol.* **19**, 171–191 (1969).
17. Nessov, L. A. New Mesozoic mammals of middle Asia and Kazakhstan and comments about evolution of theriofaunas of Cretaceous coastal plains of Asia. *Trudy Zool. Inst.* **249**, 105–133 (1993).
18. Martin, T. in *Tooth Enamel Microstructure* (eds von Koenigswald, W. & Sander, P. M.) 163–175 (Balkema, Rotterdam, 1997).
19. Luckett, W. P. in *Evolutionary Relationships among Rodents: A Multidisciplinary Analysis* (eds Luckett, W. P. & Hartenberger, J.-L.) 227–276 (Plenum, New York, 1985).
20. Lopatin, A. V. & Averianov, A. O. A new species of *Tribosphenomys* (Mammalia: Rodentiaformes) from the Paleocene of Mongolia. *New Mex. Mus. Nat. Hist. Sci. Bull.* (in the press).
21. Maddison, W. P. & Maddison, D. R. *MacClade 3.07* (Sinauer, Sunderland, Massachusetts, 1997).
22. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0b7* (Sinauer, Sunderland, Massachusetts, 2001).
23. Nessov, L. A., Archibald, J. D. & Kielan-Jaworowska, Z. Ungulate-like mammals from the Late Cretaceous of Uzbekistan and a phylogenetic analysis of Ungulatomorpha. *Bull. Carnegie Mus. Nat. Hist.* **34**, 40–88 (1998).
24. Wiblé, J. R., Rougier, G. W., Novacek, M. J. & McKenna, M. C. Earliest eutherian ear region: A petrosal referred to *Prokennalestes* from the Early Cretaceous of Mongolia. *Am. Mus. Novit.* **3322**, 1–44 (2001).
25. Li, C. K. & Ting, S. Y. in *New Cranial and Postcranial Evidence for the Affinities of the Euryomyliids (Rodentia) and Mimotomids (Lagomorpha)* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 151–158 (Springer, New York, 1993).
26. Archibald, J. D. in *Tertiary Mammals of North America*. Vol. 1. *Terrestrial Carnivores, Ungulates, and Ungulatelike Mammals* (eds C. Janis, C., Scott, K. & Jacobs, L.) 292–331 (Cambridge Univ. Press, Cambridge, 1998).
27. Novacek, M. J. *et al.* Epipubic bones in eutherian mammals from the Late Cretaceous of Mongolia. *Nature* **389**, 483–486 (1997).
28. Vaughan, T. A., Ryan, J. M. & Czaplewski, N. J. *Mammalogy* 4th edn 73 (Saunders, Fort Worth, Texas, 2000).
29. Szalay, F. S. *Evolutionary History of the Marsupials and an Analysis of Osteological Characters* 35–58, **362** (Cambridge Univ. Press, Cambridge, 1994).
30. Marshall, L. G., Case, J. A. & Woodburne, M. O. Phylogenetic relationships of the families of marsupials. *Curr. Mamm.* **2**, 433–505 (1990).
31. Archibald, J. D. & Averianov, A. O. Description of new material and revision of the Late Cretaceous mammal *Kulbeckia* (Eutheria, Zalambdalestidae). *J. Vert. Paleontol.* (in review).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The continued cooperation of the Zoological Institute, National Academy of Sciences of Uzbekistan, and particularly D. A. Azimov and Y. Chikin, is much appreciated. We thank the URBAC expedition members A. Abramov, I. Danilov, C. King, N. Morris, A. Resvyi, C. Skrabec, P. Skutschas, H.-D. Sues and D. Ward for their field help and scientific expertise. We also thank S. Azadov, A. Khodjaev, A. Salikhbaev, V. Savin, O. I. Tsaruk, B. G. Veretennikov, N. I. Kuchersky, V. V. Novikov, V. V. Poverennov, A. Prokhorenko and N. I. Pronin. We thank M. Novacek, G. Rougier and J. Wiblé for sharing their ideas, unpublished results and specimens. A. Berta, R. Etheridge and T. Reeder commented on the manuscript and T. Reeder helped with the analyses. The financial support of the National Geographic Society, the National Science Foundation, the Navoi Mining and Metallurgy Combinat, and the San Diego State University International Programs is gratefully acknowledged.

Correspondence and requests for materials should be addressed to J.D.A. (e-mail: darchibald@sunstroke.sdsu.edu).

Rapid responses of British butterflies to opposing forces of climate and habitat change

M. S. Warren*, J. K. Hill†‡, J. A. Thomas§, J. Asher*, R. Fox*, B. Huntley‡, D. B. Roy||, M. G. Telfer||, S. Jeffcoate*, P. Harding||, G. Jeffcoate*, S. G. Willis‡, J. N. Greatorex-Davies||, D. Moss|| & C. D. Thomas¶

* Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset BH20 5QP, UK

† Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

‡ Environmental Research Centre, Department of Biological Sciences, University of Durham, Durham DH1 3LE, UK

§ Centre for Ecology and Hydrology, Dorset Laboratory, Winfrith Technology Centre, Dorchester, DT2 8ZD, UK

|| Centre for Ecology and Hydrology, Monks Wood, Abbots Ripton, Huntingdon, Cambridgeshire PE28 2LS, UK

¶ Centre for Biodiversity and Conservation, School of Biology, University of Leeds, Leeds LS2 9JT, UK

Habitat degradation and climate change are thought to be altering the distributions and abundances of animals and plants throughout the world, but their combined impacts have not been assessed for any species assemblage^{1–4}. Here we evaluated changes in the distribution sizes and abundances of 46 species of butterflies that approach their northern climatic range margins in Britain—where changes in climate and habitat are opposing forces. These insects might be expected to have responded positively to climate warming over the past 30 years, yet three-quarters of them declined: negative responses to habitat loss have outweighed positive responses to climate warming. Half of the species that were mobile and habitat generalists increased their distribution sites over this period (consistent with a climate explanation), whereas the other generalists and 89% of the habitat specialists declined in distribution size (consistent with habitat limitation). Changes in population abundances closely matched changes in distributions. The dual forces of habitat modification and climate change are likely to cause specialists to decline, leaving biological communities with reduced numbers of species and dominated by mobile and widespread habitat generalists.

We studied all 46 non-migratory British butterfly species that reach their northern margins in Britain, where the summer–spring climate has warmed by approximately 1–1.5°C in the past 25 years^{5,6}. Many of these butterflies are restricted to warm local environments in Britain and have faster larval growth rates, earlier flight periods and increased abundances at higher temperatures (within the British temperature range)^{1,7–9}. Range expansions have

also been observed at northern margins^{10–13}. Under the hypothesis that climate is limiting, most species should have benefited from climate warming. In contrast, intensification of agriculture in Britain has led to 70% (range 40–97%) losses of semi-natural habitats since 1940 (ref. 13). On the basis of habitat alone, most species should have declined¹³ (see Fig. 1). Here we report how butterflies differing in dispersal and habitat specialization^{14,15} have responded to an improving climate but continued habitat degradation.

For each species, distribution change was measured as the difference in the number of 10-km grid squares occupied between 1970–82 and 1995–99 (refs 13, 16; the 1995–99 data subsampled to equalize recorder effort; see Methods), divided by the 1970–82 distribution size. Three-quarters (34/46) of species declined in distribution area. Habitat specialists fared worse than wider-countryside generalists: 26/28 specialists declined compared to 9/18 wider-countryside species (Fig. 2; $t = 2.97$, degrees of freedom, d.f., 8, $P = 0.018$). Sedentary species fared worse than mobile species: 24/26 sedentary species declined compared to 11/20 mobile species ($t = 3.68$, d.f., 10, $P = 0.004$). The effects of mobility and habitat specificity cannot be separated, as these traits are highly correlated among butterflies (26/28 specialist species are sedentary, 18/18 wider-countryside species are mobile). In conjunction, low mobility and high habitat specialization restrict species in fragmented habitats, and limit expansion across patchy, human-modified landscapes^{17–21}. Even expanding species have expanded more slowly in areas where there is less suitable habitat for them to expand into^{2,18}.

Changes in abundance were measured from weekly counts of adults at around 120 fixed sites between 1976 and 2000 (ref. 14), and summarized as the slope of the regression between log of the abundance and the year. Mobile generalists tended to increase

relative to sedentary specialists (Fig. 3; specialists versus wider-countryside $t = 2.98$, d.f., 5, $P = 0.031$; sedentary versus mobile $t = 2.93$, d.f., 4, $P = 0.043$). Changes in abundance and changes in distribution size were correlated across all species for which abundance data were available (Fig. 3; $r^2 = 0.63$, $F_{1,24} = 41.55$ (1 and 24 d.f.), $P < 0.0001$). Many species face both declining abundance and declining distribution.

The negative intercept of the regression (Fig. 3; non-phylogenetic $c = -0.148$, standard error of the mean, s.e.m. = 0.032, $P < 0.001$) suggests that abundance changes have been more favourable than distribution changes: on average, a species that showed no abundance change on monitored sites over the past 25 years declined by 15% in distribution area. The effect was stronger for sedentary specialists (21–22% distribution decline for no change in local abundance) than for mobile generalists (9–9.5% decline). In an analysis of covariance (ANCOVA) for the mobility effect, $F_{1,24} = 5.16$, $P = 0.032$; for the specialization effect, $F_{1,24} = 3.95$, $P = 0.058$.

These analyses suggest that most sedentary and specialized species, and at least half of the other species, are limited by factors other than climate. We studied this by fitting climate response surface models^{12,22} at 50-km grid resolution to the distributions of British butterflies throughout Europe²³ to estimate the extent of climatically suitable areas potentially available for each species in Britain. Generally there were good fits between observed and

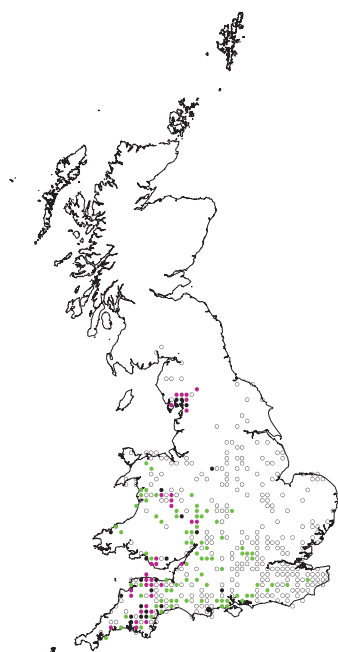


Figure 1 Reduced distribution of the high-brown fritillary. Distribution changes between 1970–82 and 1995–99 (full data set) of the high-brown fritillary *Argynnis adippe*, the habitat specialist that has shown the most rapid decline in distribution size. Black circles (10-km grid resolution) show butterfly records for populations in both 1970–82 (ref. 16) and 1995–99 (ref. 13); green shows apparent extinction (recorded 1970–82; not recorded 1995–99); pink shows new records (no record 1970–82; record in 1995–99); white circles show pre-1970 extinctions (data 1857–1969). Recording was incomplete in the past, so the real decline has been ever steeper.

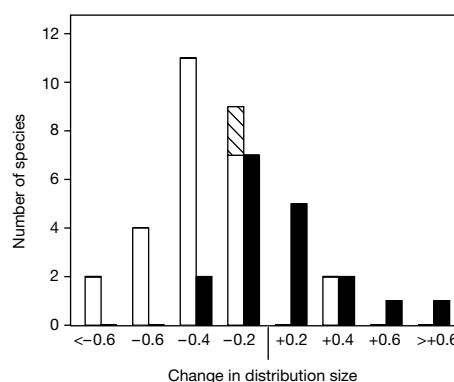


Figure 2 Proportional changes in distribution sizes of butterflies between 1970–82 and 1995–99. Sedentary specialists (white), mobile specialists (hatched) and mobile wider-countryside species (black) are shown.

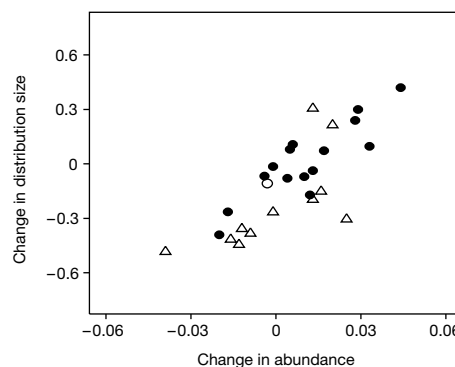


Figure 3 Correlation between changes in the abundance and distribution of butterflies. The data show the relationship between the trend in population abundance between 1976 and 2000 and the change in distribution size between 1970–82 and 1995–99, for sedentary specialists (open triangles), mobile wider-countryside species (filled circles), and one mobile specialist (open circle).

simulated distributions at a European scale (kappa goodness-of-fit^{12,24}, mean is 0.79, s.e.m., 0.007, range 0.67–0.87), showing that continental range limits of all the British study species could be described by three bioclimate variables (see Methods). Goodness-of-fit was not affected by mobility ($P = 0.2$) or specificity ($P = 0.3$). We then calculated the difference between the extent of land in Britain ($\sqrt{\text{area}}$, 10-km grid resolution) deemed climatically suitable and the area occupied by each species in 1995–99. Specialists lagged behind climate more than did species of the wider countryside (specialists' mean lag is -104.55 , s.e.m., 13.40 , $n = 28$; wider-countryside mean lag is -37.03 , s.e.m., 8.48 , $n = 18$; $t = 2.50$, d.f., 8 , $P = 0.037$), and sedentary species tended to lag behind climate

more than mobile species ($t = 2.21$, d.f., 10 , $P = 0.052$). Climatically suitable areas are apparently available for colonization, but most species (especially sedentary specialists) have failed to exploit them either because they do not contain suitable breeding habitats, or because breeding habitats are out of reach.

Three thermally limited species¹³ illustrate the patterns typical of most other species. *Plebejus argus* has low mobility, principally inhabits lowland heathland, and declined in area of occupancy by 28% (Fig. 4a). It is restricted within apparently suitable climatic areas (Fig. 4d), and is limited by habitat and dispersal more than by climate. Most declining species show comparable patterns. *Pararge aegeria* is a more mobile, wider-countryside species, capable of

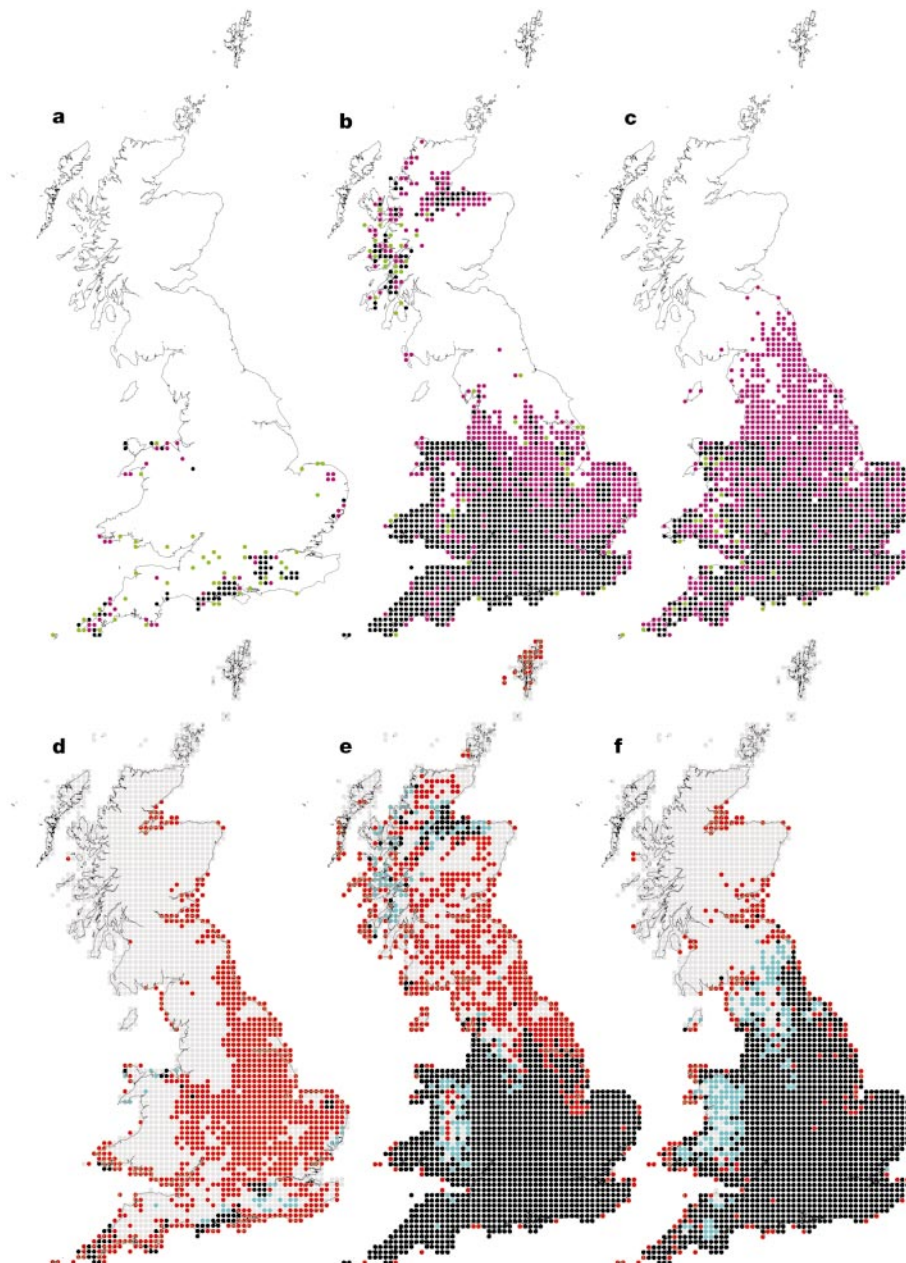


Figure 4 Climate and species ranges. The degree to which three species have changed their ranges (**a–c**, without subsampling) and are lagging behind current climates in Britain (**d–f**, 10-km grid resolution). **a, d**, Silver-studded blue, *Plebejus argus*; **b, e**, speckled wood, *Pararge aegeria*; **c, f**, comma, *Polygonia c-album*. For maps **a–c**, black circles show butterfly records for both 1970–82 (ref. 16) and 1995–99 (ref. 13); green circles show apparent extinction (recorded 1970–82; not 1995–99); pink circles show apparent

colonization (no record 1970–82; record 1995–99). For maps **d–f**, black circles (climate suitable, butterfly present) and grey circles (climate unsuitable, butterfly absent) show where observed 1995–99 and simulated distributions agree; red circles (climate predicted suitable, butterfly not recorded) and blue circles (butterfly recorded, climate deemed unsuitable) show mismatches.

inhabiting woodlands, scrub, hedgerows and shady gardens. Its distribution has expanded by 24%, but still lags behind climate¹² (Fig. 4b, e), with regional differences in its rate of expansion accurately predicted by differences in woodland cover²: both habitat and climate restrict its distribution. Most other expanding species show comparable patterns. *Polygonia c-album* is a very mobile butterfly of the wider countryside whose distribution has expanded by 30%, and shows little or no lag behind its climate (Fig. 4c, f). This species is mainly limited by climate.

Most species of non-migratory butterflies that reach the northern margins of their geographic ranges in Britain have declined over the last 30 years (as they have elsewhere in northern Europe^{21,25}) even though the climate has warmed. This is surprising because climate warming is expected to increase the range of habitats these species can inhabit^{1,19}. However, most sedentary specialists have not expanded because habitat patches are too isolated to colonize. Thus, quaternary expansions of species into areas where the climate 'improved' over the past 15,000 years³ are unlikely to be replicated for many sedentary and specialized species in heavily modified, modern landscapes⁴. In contrast, quaternary rates of decline are likely to be exceeded in regions where climate-change and habitat are both acting as agents of decline, potentially leading to serious losses of biodiversity in areas such as Mediterranean Europe^{10,25}, where most species already approach their climatic maxima. Large-scale protection and management of habitat networks are required to minimize habitat-related declines and to maximize the ability of species to track the distribution of suitable climate. If sedentary specialists continue to decline, biological communities will increasingly become dominated by mobile generalists. □

Methods

Distributions

65,826 and 437,690 separate record cards, listing all species observed on one field visit, were collated for England, Wales and Scotland for 1970–82 (ref. 16) and 1995–99 (ref. 13), respectively. To equalize recorder effort, we subsampled the 1995–99 data by randomly selecting the 1970–82 number of record cards from the 1995–99 data, subsampling separately for each 100-km Ordnance Survey grid square to retain the broad geographical distribution of 1970–82 records. Results are based on subsampling, except Figs 1 and 4 (all records) and measures of the lag between climate space and observed distributions (full 1995–99 distribution is appropriate). Statistical differences among species (according to mobility and habitat specialization) were robust when re-analysed (1) using the entire 1995–99 data set (rather than the subsample), and (2) measuring distribution change as the square root of area (in square kilometres), rather than proportional change.

Abundance

Weekly count data were collected along about 120 fixed transect routes¹⁴ and summarized as one collated index of abundance for each species in each year²⁶ (for sites where present, excluding sites after two successive zeros were recorded, including colonized sites after two successive presences). Species were included only if the abundance trend was based on data from more than 10 sites per year (mean is 36 sites per year per species, $n = 27$ species). We excluded *Polyommatus icarus* (number of generations varies regionally), *Thymelicus sylvestris* and *T. lineola* (not identified separately on transects).

Species and their characteristics

We included all southerly distributed species, except *Colias croceus*, *Pieris brassicae*, *P. rapae*, *Vanessa atalanta*, and *V. cardui*, which migrate between the UK and continental Europe, and *Papilio machaon*, which has many British records of vagrants from continental Europe.

We used a mobility ranking¹⁵ (other rankings are in close agreement), and scored species as relatively sedentary (ranks 0–2 (ref. 15), $n = 26$ species) or mobile (ranks 3–6, $n = 20$ species; ranks recalculated to exclude expansion data). On the basis of mark–release–recapture studies¹⁶, *Argynnis aglaja*, *A. paphia* and *Hipparchia semele* were re-classified as sedentary, and *Aphantopus hyperantus* as mobile. Statistical conclusions are unaffected by these changes. We scored species as wider-countryside species ($n = 18$ species) or specialists restricted to specific habitats ($n = 28$) species¹⁴.

Analyses were controlled for phylogeny²⁷ using independent phylogenetic contrasts (CAIC program²⁸). Mobility and habitat specificity were scored as binary variables. Unless stated, all analyses were phylogenetically controlled: non-phylogenetic analyses (not shown) supported all conclusions. Non-phylogenetic methods were used for analysing the regression intercepts of abundance–distribution relationships because regressions based on phylogenetic contrasts must pass through the origin²⁸.

Climate response surface models

Butterfly distributions^{23,13} were converted to presence/absence on a 50-km Universal Transverse Mercator grid (Azores to longitude 30° E; Mediterranean coast to Svalbard; 2,648 cells). We used mean monthly temperature, precipitation and cloudiness relating to the climate normal period 1931–60 (ref. 29) to interpolate values for the midpoint and mean elevation of each 50-km cell²². For each cell, we computed (1) mean temperature of the coldest month (related to overwintering survival); (2) annual temperature sum over 5 °C (development potential for immature stages); (3) an index of moisture availability (ratio of actual to potential evapotranspiration). We fitted climate response surfaces describing each species' European distribution using these variables^{12,22}.

Climate response surfaces generated at 50-km grid resolution were applied to finer-scale (10-km grid) climate data to simulate the 1961–90 extent of climatically suitable areas for each species in Britain. Bioclimate values were derived for the midpoint and mean elevation of each 10-km cell (2,805 cells) in Britain using the same techniques and data sets. For each species, we computed the difference between the area of suitable climate (black plus red circles in Fig. 4d–f) and the area currently occupied (black plus blue circles in Fig. 4d–f). Most blue circles (butterfly recorded, climate deemed unsuitable) in Fig. 4d–f are from areas of high relief where the species occurs in warm habitats (for example, south-facing slopes) below the mean elevation of the grid cell.

Received 22 June; accepted 10 September 2001.

1. Thomas, J. A., Rose, R. J., Clarke, R. T., Thomas, C. D. & Webb, N. R. Intraspecific variation in habitat availability among ectothermic animals near their climatic limits and at their centres of range. *Funct. Ecol.* **13**, 55–64 (1999).
2. Hill, J. K. *et al.* Impacts of landscape structure on butterfly range expansion. *Ecol. Lett.* **4**, 313–321 (2001).
3. Coope, G. R. in *Insects in a Changing Environment* (eds Harrington, R. & Stork, N.) 29–48 (Academic, London, 1995).
4. Groombridge, B. *Global Biodiversity* (Chapman & Hall, London, 1992).
5. Jones, P. D. & Hulme, M. in *Climates of the British Isles: Present, Past and Future* (eds Hulme, M. & Barrow, E.) 173–196 (Routledge, London, 1997).
6. Roy, D. B. & Sparks, T. H. Phenology of British butterflies and climate change. *Glob. Change Biol.* **6**, 407–416 (2000).
7. Thomas, J. A. Holocene climate changes and warm man-made refugia may explain why a sixth of British butterflies possess unnatural early-successional habitats. *Ecography* **16**, 278–284 (1993).
8. Thomas, J. A., Moss, D. & Pollard, E. Increased fluctuations by butterfly populations towards the northern margins of species' ranges. *Ecography* **17**, 215–220 (1994).
9. Roy, D. B., Rothery, P., Moss, D., Polard, E. & Thomas, J. A. Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. *J. Anim. Ecol.* **70**, 201–217 (2001).
10. Parmesan, C. *et al.* Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579–583 (1999).
11. Hughes, L. Biological consequences of global warming: is the signal already present? *Trends Ecol. Evol.* **15**, 56–61 (2000).
12. Hill, J. K., Thomas, C. D. & Huntley, B. Climate and habitat availability determine 20th century changes in a butterfly's range margins. *Proc. R. Soc. Lond. B* **266**, 1197–1206 (1999).
13. Asher, J. *et al.* *The Millennium Atlas of Butterflies in Britain and Ireland* (Oxford Univ. Press, Oxford, 2001).
14. Pollard, E. & Yates, T. J. *Monitoring Butterflies for Ecology and Conservation* (Chapman & Hall, London, 1993).
15. Dennis, R. L. H. & Shreeve, T. G. Diversity of butterflies on British Islands: ecological influences underlying the roles of area and isolation and the size of the faunal source. *Biol. J. Linn. Soc.* **60**, 257–275 (1997).
16. Heath, J., Pollard, E. & Thomas, J. A. *Atlas of Butterflies in Britain and Ireland* (Viking, Harmondsworth, 1984).
17. Thomas, C. D. Dispersal and extinction in fragmented landscapes. *Proc. R. Soc. Lond. B* **267**, 139–145 (2000).
18. Hill, J. K., Thomas, C. D., Fox, R., Moss, D. & Huntley, B. in *Insect Movement: Mechanisms and Consequences* (eds Woivod, I., Reynolds, D. & Thomas, C. D.) 415–441 (CABI, London, 2001).
19. Thomas, C. D. *et al.* Ecological and evolutionary processes at expanding range margins. *Nature* **411**, 577–581 (2001).
20. Cowley, M. J. R., Thomas, C. D., Thomas, J. A. & Warren, M. S. Flight areas of British butterflies: assessing species status and decline. *Proc. R. Soc. Lond. B* **266**, 1587–1592 (1999).
21. Maes, D. & Van Dyck, H. Butterfly diversity loss in Flanders (north Belgium): Europe's worst case scenario? *Biol. Conserv.* **99**, 263–276 (2001).
22. Huntley, B., Berry, P. M., Cramer, W. & McDonald, A. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *J. Biogeog.* **22**, 967–1001 (1995).
23. Tolman, T. *Butterflies of Britain and Europe* (HarperCollins, London, 1997).
24. Monserud, R. A. & Leemans, R. Comparing global vegetation maps with the Kappa statistic. *Ecol. Model.* **62**, 275–293 (1992).
25. Van Swaay, C. & Warren, M. S. *Red Data Book of European Butterflies (Rhopalocera)* (Nature and Environment Series No. 99, Council of Europe, Strasbourg, 1999).
26. Moss, D. & Pollard, E. Calculation of collated indices of abundance of butterflies based on monitored sites. *Ecol. Entomol.* **18**, 77–83 (1993).
27. Cowley, M. J. R. *et al.* Density–distribution relationships in British butterflies. I. The effect of mobility and spatial scale. *J. Anim. Ecol.* **70**, 410–425 (2001).
28. Purvis, A. & Rambaut, A. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comp. Appl. Biosci.* **11**, 247–251 (1995).
29. Leemans, R. & Cramer, W. Research Report RR-91-18 (International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 1991).

Acknowledgements

We thank the many thousand recorders who contributed to the atlas surveys and transect schemes. We also thank W. Cramer for providing the spline surfaces used to interpolate the climate anomalies, and P. Mayhew for giving helpful advice on using CAIC. This work was supported by the Butterfly Conservation, the Esmée Fairbairn Foundation, the Vincent Wildlife Trust, the Joint Nature Conservation Committee, the Centre for Ecology and Hydrology, and NERC.

Correspondence and requests for materials should be addressed to C.D.T. (e-mail: c.d.thomas@leeds.ac.uk).

Perceptual basis of bimanual coordination

Franz Mechsner, Dirk Kerzel, Günther Knoblich & Wolfgang Prinz

Max Planck Institute for Psychological Research, Department of Cognition and Action, Amalienstrasse 33, D-80799 Munich, Germany

Periodic bimanual movements are often the focus of studies of the basic organizational principles of human actions^{1–25}. In such movements there is a typical spontaneous tendency towards mirror symmetry. Even involuntary slips from asymmetrical movement patterns into symmetry occur, but not vice versa. Traditionally, this phenomenon has been interpreted as a tendency towards co-activation of homologous muscles, probably originating in motoric neuronal structures. Here we provide evidence contrary to this widespread assumption. We show for two prominent experimental models—bimanual finger oscillation¹ and bimanual four-finger tapping²—that the symmetry bias is actually towards spatial, perceptual symmetry, without regard to the muscles involved. We suggest that spontaneous coordination phenomena of this kind are purely perceptual in nature. In the case of a bimanual circling model, our findings reveal that highly complex, even ‘impossible’ movements can easily be performed with only simple visual feedback. A ‘motoric’ representation of the performed perceptual oscillation patterns is not necessary. Thus there is no need to translate such a ‘motoric’ into a ‘perceptual’ representation or vice versa, using ‘internal models’ (ref. 29). We suggest that voluntary movements are organized by way of a representation of the perceptual goals, whereas the corresponding motor activity, of sometimes high complexity, is spontaneously and flexibly tuned in.

How do coordinative processes in the motor system and in the domain of perception and imagery contribute to the organization of voluntary movement? Spontaneous coordination phenomena such as the symmetry tendency in bimanual movements are of particular interest here. The traditional view is that the symmetry tendency is due to a bias towards co-activation of homologous muscles^{1,3}, probably originating in motoric neuronal structures. Recently, the possible influence of perception and perceptual imagery on spontaneous coordination phenomena has been stressed^{4–8,26}. However, many of these studies tend to assume that motoric, or efferent, constraints are also of central importance. Clear experimental evidence is lacking.

In our first experiment we addressed the symmetry tendency in a classical bimanual finger oscillation model^{1,2,9,10}: a person stretches out both index fingers and oscillates them in mirror symmetry or in parallel (Fig. 1a, b). The symmetrical mode is much more stable than the parallel mode. With increasing oscillation frequencies, a parallel pattern often involuntarily switches into a mirror-symmetrical movement pattern. In contrast, symmetrical movements never switch into asymmetry. Is this symmetry bias

towards co-activation of homologous muscles or towards perceptual, spatial symmetry?

Participants ($n = 8$) performed bimanual index-finger oscillations, either in symmetry or in parallel, with both movement instructions (symmetry or parallelity) defined in visual, perceptual space. To register trajectory, both fingers were inserted in cuffs of 50-g weight, with a graphics tablet stylus attached to each finger. The hands were individually put either palm up or palm down. Thus, there were four bimanual hand positions (Fig. 1c–f). If both palms are either up or down, the hand position is congruous. If one palm is up and the other is down the hand position is incongruous. In a session, each combination of movement instruction and hand position was performed four times, in a total of 32 randomized trials. In a trial, a metronome pulse paced the oscillation frequency from 1.4 Hz up to 3.6 Hz, in a time interval of 24 s. Participants were requested to execute one full movement cycle on each beat. Should the movement pattern change, participants were instructed to give in and perform the more comfortable pattern¹¹.

The experimental rationale, as adopted from designs in the literature^{7,12–14}, was as follows. With a congruous hand position, perceptual movement symmetry goes along with periodic co-activation of homologous muscles. Thus, a bias towards symmetrical oscillation is to be expected, as it is a replication of results reported previously. The critical condition is with incongruous hand position, because perceptual parallelity goes along with co-activation of homologous muscles. Thus, if there is a dominant tendency towards co-activation of homologous muscles, a bias towards parallel oscillation is to be expected. If there is a dominant tendency towards perceptual symmetry, there should be a bias towards symmetrical oscillation.

The results were clear: independent of hand position, an instructed symmetrical oscillation pattern is always stable, whereas instructed parallel oscillations tend to disintegrate and to switch into symmetry. Figure 2 demonstrates this by showing histograms of the relative phase of the fingertips, as defined in a position versus velocity coordinate system^{11,15}. Zero degrees relative phase means symmetry, whereas 180° relative phase means parallelity, in perceptual space. Relative phase was calculated on every right reversal of the left finger.

We define a relative phase of $0 \pm 60^\circ$ as symmetry, of $180 \pm 60^\circ$ as

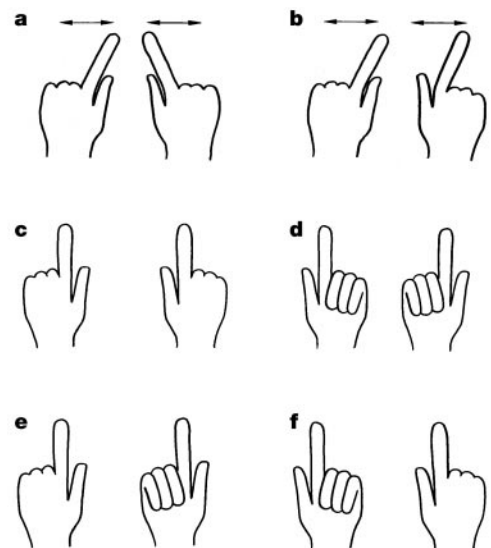


Figure 1 Instructed, synchronous finger oscillation patterns and hand positions.

a, Symmetrical movement. **b**, Parallel movement. **c**, **d**, Congruous positions with both palms up or both palms down. **e**, **f**, Incongruous positions with one palm up and the other palm down.

parallelity and of $90 \pm 30^\circ$ or $270 \pm 30^\circ$ as an intermediate mode¹⁶. The distribution of these modes was calculated for the last 4 s in each trial. With the incongruous hand positions, the proportion of other modes than the instructed one was significantly higher under an instruction of parallelity (59.17%) than under symmetry instruction (9.70%) ($P < 0.001$, paired t -test). With the incongruous hand positions and parallelity instruction, symmetry was the most frequent coordination mode among the non-instructed modes (symmetry, 46.38%; intermediate, 12.79%, $P < 0.001$). With the incongruous hand positions and symmetry instruction, there were virtually no transitions into parallelity (parallelity, 0.75%). This general pattern of results does not change when a view of the fingers is prevented and participants have to rely solely on proprioception (data not shown).

We conclude that the symmetry tendency in the bimanual finger oscillation model is a tendency towards perceptual, spatial symmetry, rather than towards co-activation of homologous muscles.

This is a challenging result. The idea that constraints in the motor system might bring about the symmetry tendency has been based on the assumption that there is a general tendency towards co-activation of homologous muscles. Theorists proposed, for example, that mechanisms such as bilateral cross-talk in motoric neuronal

structures might cause this tendency¹⁷. Or there might be a mechanism that eases motor programming by taking advantage of homologies¹⁸. If it turns out that there is no general tendency towards co-activation of homologous muscles, the very existence of such motoric constraints may also seriously be doubted.

In our second experiment we addressed the symmetry tendency in a bimanual finger-tapping model². A person stretches out the index (I) and middle (M) fingers of both hands and taps on the table with the fingertips. We denote this finger combination by (MI·IM). Symmetrical tapping is defined as synchronous tapping of both index fingers in periodic alternation to synchronous tapping of both middle fingers, that is (I·I·), (M·M·), (I·I·), and so on. Parallel tapping means synchronous tapping of the left middle and the right index finger in periodic alternation to synchronous tapping of the left index and the right middle finger: (M·I·), (I·M·), and so on. In parallel tapping, the movement pattern is less stable than in symmetrical tapping. Involuntary transitions into symmetry occur with higher tapping frequencies. Is this symmetry bias towards co-activation of homologous muscles, and thus possibly motoric in nature, or towards perceptual, spatial symmetry?

Participants ($n = 10$) were instructed to perform symmetrical as well as parallel bimanual tapping patterns using four fingers. For

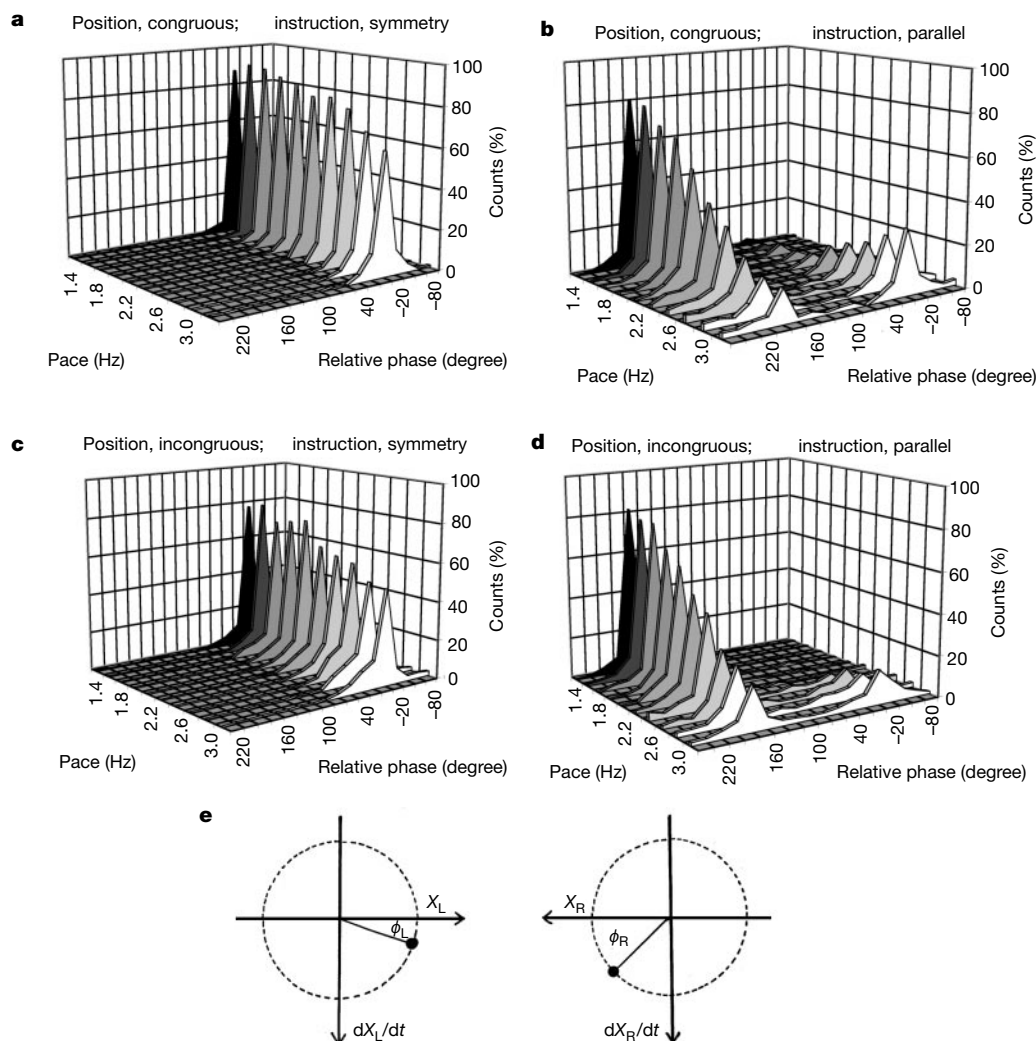


Figure 2 Relative phase of the fingertips averaged across subjects in experiment 1. **a**, Congruous hand positions and symmetrical movement instruction. **b**, Congruous hand positions and parallel movement instruction. **c**, Incongruous hand positions and symmetrical movement instruction. **d**, Incongruous hand

positions and parallel movement instruction. **e**, Coordinate systems as used for defining relative phase, Θ (ref. 15). X_R and X_L indicate the transversal positions of the right and left index finger, respectively; dX_R/dt and dX_L/dt their velocities. $\Theta = \phi_R - \phi_L = \tan^{-1}[(dX_R/dt)/X_R] - \tan^{-1}[(dX_L/dt)/X_L]$.

registration of the onset time of the taps, the fingers tapped on four small metal plates, which were connected to a computer. The acting fingers of an individual hand were either the index and middle finger, or the middle and ring finger. Thus, there were four bimanual finger settings. If the finger combinations of both hands are identical, the finger setting is called congruous, otherwise the finger setting is incongruous. For example, symmetrical tapping in an (MI-MR) finger setting means ($_I_M_$), ($M_ _R$), ($_I_M_$),... In a session, each combination of movement instruction and finger setting was performed four times, in a total of 32 randomized trials. In a trial, a metronome pulse paced the tapping frequency from 1 Hz up to 3 Hz, in a time interval of 45 s. Participants were requested to execute one full movement cycle on each beat. Should the movement pattern change, participants were instructed to give in and perform the more comfortable pattern.

The experimental rationale was analogous to that of experiment 1: with a congruous finger setting, perceptual movement symmetry goes along with periodic co-activation of homologous fingers and muscles. Thus a bias towards symmetry is to be expected, as it is a replication of results reported previously. The critical condition is, again, the incongruous setting. In this case, homologous fingers are co-active only in the parallel pattern. In the case of a tendency towards co-activation of homologous muscles, the parallel pattern should be the more stable one. In the case of a bias towards perceptual symmetry, the symmetrical pattern should be of greater stability.

The results were, again, clear. Independent of the finger setting, an instructed symmetrical pattern was always stable whereas parallel patterns heavily switched into symmetry with increasing tapping frequencies. Figure 3 demonstrates this by displaying the percentage of symmetrical taps for congruous and incongruous finger settings, under both instructions.

Where one finger of the right hand and one finger of the left hand tap together in a time window of 70 ms, this event is categorized as a synchronous tap. Each synchronous tap can individually be categorized as symmetrical or parallel. Symmetrical and parallel taps were counted separately. For analysis, each trial was separated into four intervals of equal length, 5–15 s (mean frequency 1.1 Hz), 15–25 s (1.4 Hz), 25–35 s (1.9 Hz) and 35–45 s (2.9 Hz). To statistically confirm the reported results the data were entered into a $2 \times 2 \times 4$ repeated-measurements analysis of variance; the factors were instruction, finger setting, and time interval. An α -level of 0.01 was always applied. The analysis revealed highly significant main effects for instruction ($F_{1,9} = 30$) and for interval ($F_{3,27} = 14$), as well as a highly significant interaction between these two factors ($F_{3,27} = 16$). The main effect for congruency and the remaining interactions were not significant. This general pattern of results does not change

when the fingers cannot be seen and participants have to rely solely on proprioception (data not shown).

We conclude that the symmetry tendency in the investigated model of four-finger tapping is not a tendency towards co-activation of homologous muscles, but instead arises in the functional domain of perception and perceptual imagery.

Observations on further bimanual oscillation models confirm the obtained results. There is obviously no general tendency towards co-activation of homologous muscles. Therefore, motoric interpretations of the symmetry tendency are no longer plausible. Instead, the hypothesis that the symmetry tendency in bimanual oscillation is purely perceptual in nature seems justified.

We speculate that perception and imagery provide a plausible, unifying explanatory principle, not only for the symmetry tendency, but also for other spontaneous bimanual coordination phenomena of a similar kind.

In experiments 1 and 2, the perceptual oscillation pattern—that is, symmetry or parallelity (antiphase)—could always be predicted from the bimanual muscle activation pattern. This includes the possibility that the central nervous system might bring about the intended movement pattern by way of activating a corresponding motoric pattern. It is a common assumption in theories concerning human motor control that voluntary movement performance relies on a translation of a characteristic, corresponding motoric pattern into the intended sensory pattern, and vice versa^{27–29}. In our next experiment we investigated whether the existence of such characteristic, corresponding motoric patterns is necessary for symmetrical and antiphase bimanual circling.

In our third experiment we addressed a bimanual circling model, as introduced previously (refs 19, 20). When humans circle both hands in a horizontal plane, only a few synchronous circling patterns are stable. There is a tendency towards synchronization and, in particular, symmetry. Especially difficult, and almost impossible for untrained persons, are circling patterns under non-harmonic frequency ratios of the hands, such as a 5:4 or 4:3 frequency ratio.

Right-handed participants ($n = 8$) circled two visible flags, by way of two cranks that were hidden under the table (Fig. 4a). The left flag circled directly above the left crank and hand, whereas the right flag circled in a 4:3 frequency ratio to the right crank and hand, owing to a gear system. Thus, iso-frequency in the flags went together with a 4:3 frequency ratio in the cranks (hands). Movements were registered by way of a graphics stylus tablet inserted in each crank.

The participants either circled both flags (hand agency), or circled the right flag alone, while a motor driven by the experimenter (motor agency) circled the left flag. Participants were instructed to

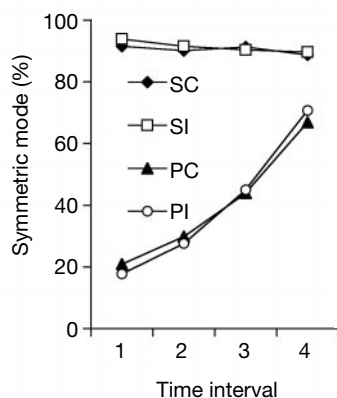


Figure 3 Percentage of symmetrical taps, in experiment 2, split by movement instruction and bimanual finger setting (see text). SC, instruction is symmetry; finger setting is congruous. SI, symmetry; incongruous. PC, parallelity; congruous. PI, parallelity; incongruous.

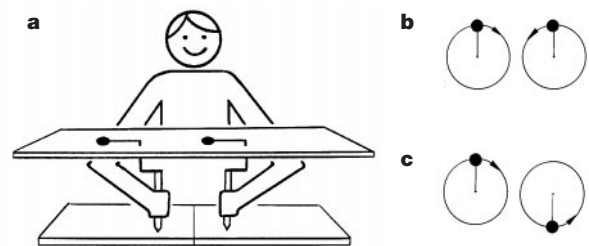


Figure 4 Apparatus used in experiment 3, and instructed synchronous circling patterns of the flags. **a**, Apparatus. The participant circles two visible flags using his or her hidden hands. The left flag moves coincidentally with the left hand whereas the right flag moves according to a well defined angle and/or frequency transformation with regard to the right hand (see text). **b**, Symmetry, that is, 0° relative angle. **c**, Antiphase, that is, 180° relative angle. Relative angle is defined in analogy to relative phase as described in Fig. 2e, this time in Y versus X position coordinate systems instead of dX/dt versus X .

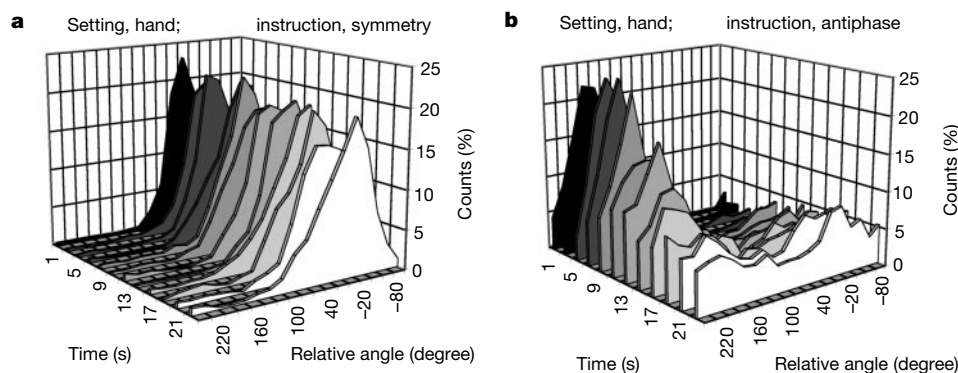


Figure 5 Histograms of relative angle of the circling flags, averaged across subjects in experiment 3. The right flag circled in a 4:3 frequency ratio to the right hand (see text).

circle the flags inwards and maintain the visual circling pattern either in mirror symmetry (0° relative angle, see Fig. 4b) or in antiphase (180° relative angle, see Fig. 4c). In a trial of 30-s duration, participants began at a slow, comfortable velocity and then speeded up to a velocity they considered fast, but not beyond the point they lost visual control. Before the experiment, participants were told that in situations where they felt a tendency towards a circling mode other than instructed, they should give in and maintain the preferred mode. After a 15–20 min training period, each participant performed 40 experimental trials, in a blocked 2 agency $\times$ 2 movement instruction design, which was balanced across subjects.

The main rationale was as follows. First, owing to the frequency transformation in the right hand and flag system, symmetry or antiphase in the flags cannot be predicted from the corresponding hand movement pattern, which is identical in both cases. Second, performing bimanual oscillations in a 4:3 frequency ratio is practically impossible for naive subjects as an explicit task². This was confirmed by a control experiment that was adapted to our model. In consequence, no body-oriented strategy is possible to bring about iso-frequency in the flags, let alone symmetry or antiphase. If participants were able to perform any of the instructed movement patterns, this would certainly be solely due to visual strategies. It would be of additional interest whether manifest transitions from antiphase into symmetry occur in the flags, as such a tendency is certainly perceptual in nature.

All participants were able to circle the flags in symmetry and speed up, under both hand and motor agency. Antiphase was also manageable but was corrupted with increasing circling velocities. In addition, switches into symmetry occurred. Figure 5 demonstrates this by showing histograms of the relative angle of the flags. A relative angle of 0° means symmetry, whereas 180° means antiphase. Relative angle was calculated on every right reversal of the left flag.

We defined coordination modes in analogy to the procedure reported for experiment 1, and calculated their distribution for the last 4 s of the trials. In the last 4 s, under both agencies, the proportion of modes in the flags other than the instructed one was significantly higher under antiphase instruction than under symmetry instruction (hand agency, 65.38% versus 21.03%, respectively; motor agency, 71.31% versus 13.44%, respectively, in both cases, $P < 0.001$). Under antiphase instruction, symmetry tends to be the most frequent coordination mode among the non-instructed modes (hand agency: symmetry, 38.61%; intermediate, 26.77%, $P = 0.076$; motor agency: symmetry, 50.79%; intermediate, 20.52%, $P < 0.001$). A certain symmetry tendency with regard to the hidden hands was also revealed, but solely under hand agency and symmetry instruction. Anecdotal evidence seems to suggest that attention to the hands disrupts control of iso-frequency in the flags.

a, Hand agency and symmetry movement instruction. **b**, Hand agency and antiphase movement instruction.

We conclude that symmetry and antiphase in the flags can be achieved in visual space even though there is no specific translation of characteristic body activity patterns into these characteristic perceptual patterns. To bring about the instructed flag movements, participants easily perform otherwise impossible body movements. Finally, a tendency to circle the flags in symmetry, independent of what the hands are doing, supports the notion that the symmetry tendency in the bimanual circling model is purely perceptual in nature.

Taken together, our results provide evidence that bimanual coordination is much more independent of coordinative processes in the motor system than is often thought. The symmetry tendency in bimanual movements is independent of muscular and motoric constraints and is thus purely perceptual in nature. A perceptual interpretation seems to hold also for other spontaneous coordination phenomena of the same kind. In our last experiment, not only spontaneous but also intentional symmetry and antiphase are clearly organized exclusively in the domain of perception and perceptual imagery. There is no need of a representational counterpart in the motor system and thus, trivially, no need of its translation using 'internal models' (ref. 29). This is contrary to widespread assumptions concerning human movement organization, as mentioned above. We speculate that voluntary movements are, in general, organized by way of a simple representation of the perceptual goals³⁰, whereas the corresponding motor activity of, sometimes extreme, formal complexity is spontaneously tuned in. It may be this kind of movement organization that makes the richness and complexity of human voluntary movements possible, be it in sports and dance, skilful tool use, or language. □

Received 18 June; accepted 07 September 2001.

1. Kelso, J. A. S. Phase transitions and critical behavior in human bimanual coordination. *Am. J. Physiol. Regul.* **15**, R1000–R1004 (1984).
2. Kelso, J. A. S. *Dynamic patterns: The Self-organization of Brain and Behavior* (MIT Press, Cambridge, Massachusetts, 1995).
3. Johnson, K. A. et al. Bimanual co-ordination in Parkinson's disease. *Brain* **121**, 743–753 (1998).
4. Kelso, J. A. S. The informational character of self-organized coordination dynamics. *Hum. Mov. Sci.* **13**, 393–413 (1994).
5. Swinnen, P. S., Jardin, K., Meulenbroek, R., Dounskaia, N. & Hofkens-Van Den Brandt, M. Egocentric and allocentric constraints in the expression of patterns of interlimb coordination. *J. Cogn. Neurosci.* **9**, 348–377 (1997).
6. Swinnen, P. S. et al. Exploring interlimb constraints during bimanual graphic performance: effects of muscle grouping and direction. *Behav. Brain Res.* **90**, 79–87 (1998).
7. Buchanan, J. J. & Kelso, J. A. S. Posturally induced transitions in rhythmic multijoint limb movements. *Exp. Brain Res.* **94**, 131–142 (1993).
8. Kelso, J. A. S., Fink, P. W., DeLaplaine, C. R. & Carson, R. G. Haptic information stabilizes and destabilizes coordination dynamics. *Proc. R. Soc. Lond. B* **268**, 1207–1213 (2001).
9. Haken, H., Kelso, J. A. S. & Bunz, H. A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* **51**, 347–356 (1985).
10. Haken, H. *Principles of Brain Functioning* (Springer, Berlin, 1996).
11. Lee, T. D., Blandin, Y. & Proteau, L. Effects of task instructions and oscillation frequency on bimanual coordination. *Psychol. Res.* **59**, 100–106 (1996).
12. Baldissera, F., Cavallari, P. & Civaschi, P. Preferential coupling between voluntary movements of ipsilateral limbs. *Neurosci. Lett.* **34**, 95–100 (1982).

13. Kelso, J. A. S., Buchanan, J. J. & Wallace, S. A. Order parameters for the neural organization of single, multi-joint limb movement patterns. *Exp. Brain Res.* **85**, 432–444 (1991).
14. Riek, S., Carson, R. G. & Byblow, W. D. Spatial and muscular dependencies in bimanual coordination. *J. Hum. Mov. Stud.* **23**, 251–265 (1992).
15. Scholz, J. P. & Kelso, J. A. S. A quantitative approach to understanding the formation and change of coordinated movement patterns. *J. Mot. Behav.* **21**, 122–144 (1989).
16. Carson, R. G. The dynamics of isometric bimanual coordination. *Exp. Brain Res.* **105**, 465–476 (1995).
17. Cattaert, D., Semjen, A. & Summers, J. J. Simulating a neural cross-talk model for between-hand interference during bimanual circle drawing. *Biol. Cybern.* **81**, 343–358 (1999).
18. Heuer, H. Structural constraints on bimanual movements. *Psychol. Res.* **55**, 83–98 (1993).
19. Semjen, A., Summers, J. J. & Cattaert, D. The coordination of the hands in bimanual circle drawing. *J. Exp. Psychol. Hum. Percept.* **21**, 1139–1157 (1995).
20. Carson, R. G., Thomas, J., Summers, J. J., Walters, M. R. & Semjen, A. The dynamics of bimanual circle drawing. *Q. J. Exp. Psychol.* **50A**, 664–683 (1997).
21. Sternad, D. Debates in dynamics: a dynamical systems perspective on action and perception. *Hum. Mov. Sci.* **19**, 407–423 (2000).
22. Riley, M. A. & Turvey, M. T. Dynamics in action: intentional behavior as a complex system. *Am. J. Psychol.* **114**, 160–169 (2001).
23. Park, H., Collins, D. R. & Turvey, M. T. Dissociation of muscular and spatial constraints on patterns of interlimb coordination. *J. Exp. Psychol. Hum.* **27**, 32–47 (2001).
24. Saltzman, E. L. in *Mind as Motion. Explorations in the Dynamics of Cognition* (eds Port, R. F. & van Gelder, T.) 149–173 (MIT Press, Cambridge, 1995).
25. Amazeen, P. G., Amazeen, E. L. & Turvey, M. T. Breaking the reflectional symmetry of interlimb coordination dynamics. *J. Mot. Behav.* **30** (3), 199–216 (1998).
26. Zaal, F. T. J. M., Bingham, G. P. & Schmidt, R. C. Visual perception of mean relative phase and phase variability. *J. Exp. Psychol. Hum. Percept.* **26**, 1209–1220 (2000).
27. Schmidt, R. A. in *Human Motor Behavior: An Introduction* (ed. Kelso, J. A. S.) 189–235 (Erlbaum, Hillsdale, New Jersey, 1982).
28. Jordan, M. I. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 597–609 (MIT Press, Cambridge, 1995).
29. Wolpert, D. M. & Ghahramani, Z. Computational principles of movement neuroscience. *Nature Neurosci. (Suppl.)* **3**, 1212–1217 (2000).
30. Prinz, W. Perception and action planning. *Europ. J. Cogn. Psychol.* **9** (20), 129–154 (1997).

Acknowledgements

We wish to thank S. Jordan for discussions; F. Banci for constructing the apparatus used in experiment 3; S. Alessio, B. Schroer and M. Hove for running the experiments; and S. Hass for suggestions concerning the experimental procedure.

Correspondence and requests for materials should be addressed to E.M. (e-mail: mechsner@mpipf-muenchen.mpg.de).

Energetics of ion conduction through the K⁺ channel

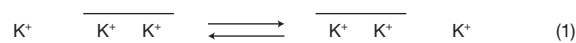
Simon Bernèche & Benoît Roux

Department of Biochemistry, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA; Département de Physique, Université de Montréal, Montréal, Québec H3C 3J7, Canada

K⁺ channels are transmembrane proteins that are essential for the transmission of nerve impulses. The ability of these proteins to conduct K⁺ ions at levels near the limit of diffusion is traditionally described in terms of concerted mechanisms in which ion-channel attraction and ion–ion repulsion have compensating effects, as several ions are moving simultaneously in single file through the narrow pore^{1–4}. The efficiency of such a mechanism, however, relies on a delicate energy balance—the strong ion-channel attraction must be perfectly counterbalanced by the electrostatic ion–ion repulsion. To elucidate the mechanism of ion conduction at the atomic level, we performed molecular dynamics free energy simulations on the basis of the X-ray structure of the KcsA K⁺ channel⁴. Here we find that ion conduction involves transitions between two main states, with two and three K⁺ ions occupying the selectivity filter, respectively; this process is reminiscent of the ‘knock-on’ mechanism proposed by Hodgkin and Keynes in 1955¹. The largest free energy barrier is on the order of 2–3 kcal mol^{−1}, implying that the process of ion conduction is limited by diffusion. Ion–ion repulsion, although essential for

rapid conduction, is shown to act only at very short distances. The calculations show also that the rapidly conducting pore is selective.

The crystallographic structure of the KcsA K⁺ channel revealed that the pore comprises a wide, nonpolar cavity of 8 Å radius on the intracellular side, leading up on the extracellular side to a narrow pore of 12 Å that is lined exclusively by main chain carbonyl oxygens⁴. This region of the pore acts as a ‘selectivity filter’ by allowing only the passage of K⁺ ions across the cell membrane⁴, whereas the wide cavity helps overcome the dielectric barrier caused by the cell membrane⁵. The translocation of K⁺ ions in single file through the narrowest region of the pore is expected to be the rate-limiting step in the conduction mechanism. This process can be represented schematically:



in which the approach of one ion from one side of the selectivity filter is coupled to the simultaneous exit of an other ion on the opposite side. Although such a concerted mechanism is consistent with long-held views of ion conduction through K⁺ channels^{1–4}, how it takes place at the atomic level remains unresolved. A simple calculation shows that the direct ion–ion repulsion varies by tens of kcal mol^{−1} when two or three ions are in the pore. Somehow, the K⁺ channel is able to exploit such large energies in a productive manner to yield a flux of about 10⁸ ions s^{−1}. This implies that there is no significant activation free energy barrier opposing the concerted ion translocation. How can this be possible?

Although the available experimental data provide a wealth of information about the structure and function of K⁺ channels, theoretical considerations are necessary for understanding the energetics of ion conduction at the atomic level. One approach to refine our understanding of complex biomolecular systems is to use

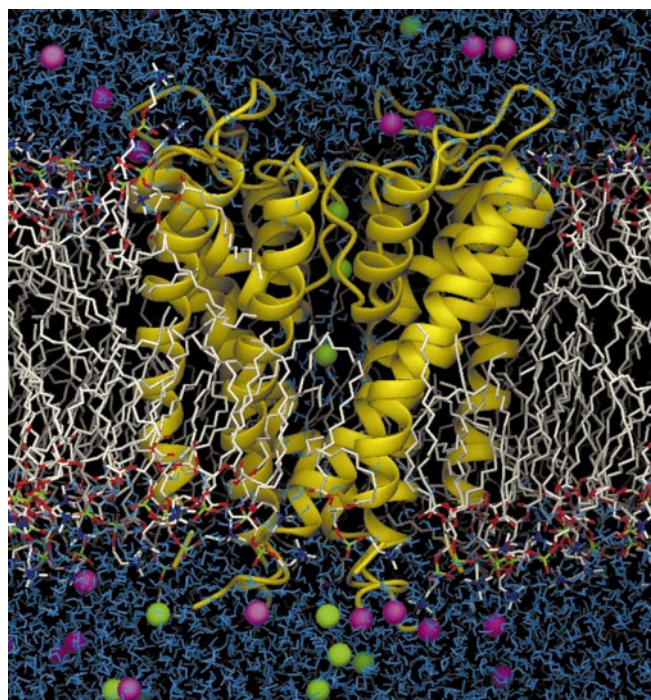


Figure 1 Molecular representation of the atomic model of the KcsA K⁺ channel embedded in an explicit DPPC phospholipid membrane bathed by a 150 mM KCl aqueous salt solution¹¹.

molecular dynamics simulations⁶. However, straight molecular dynamics trajectories of KcsA (although informative) remain limited because the average time required for the permeation of a single ion is typically much longer than can be simulated currently^{7–11}. To circumvent these difficulties and characterize the mechanism of ion conduction quantitatively, we have computed the multi-ion free energy profile governing the elementary microscopic steps of ion translocation in the pore. The free energy profile or potential of mean force (PMF)¹² is the dominant factor governing the mechanism of ion conduction¹³. The free energy surface for ion conduction was calculated using umbrella sampling simulations¹⁴, a powerful computational approach to characterize complex biomolecular systems quantitatively. The simulation system is shown in Fig. 1. First, molecular dynamics simulations are performed in the presence of an artificial biasing potential between the ions and the selectivity filter to obtain a better sampling of the relevant configurations of the system. Next, the bias introduced by this potential is rigorously removed after analysis to characterize the unbiased free energy surface of the system.

The results of the calculations are shown in Fig. 2. To aid visualization the results are presented as two-dimensional topographic maps of the free energy landscape governing the ion conduction through the selectivity filter of the K⁺ channel. To produce the two-dimensional maps, the full free energy function $W(Z_1, Z_2, Z_3)$, which depends on the position of the three ions along the channel axis, has been projected onto two different planes with reduced reaction coordinates—the ions are numbered from 1 to 3 starting from the extracellular side. In the first free energy map (Fig. 2, left), the reduced PMF is shown as a function of the position

of the ion in the cavity, Z_3 , and the position of the centre-of-mass of the two ions in the selectivity filter, Z_{12} . In the second free energy map (Fig. 2, right), the reduced PMF is shown as a function of the position of the outermost ion near the extracellular end of the pore, Z_1 , and the position of the centre-of-mass of the two ions in the selectivity filter, Z_{23} (see Methods for further details). Using such reduced reaction coordinates is meaningful because the motions of two ions located in the narrow selectivity filter are highly correlated¹¹. Typical configuration of the selectivity filter with the three K⁺ permeating ions and nearby water molecules (*a–f*) are also given in Fig. 2 to show the elementary steps of ion conduction. Figure 2, with the two-dimensional maps and the associated configurations (*a–f*), provides essentially the information connecting the left and right sides of Scheme 1 at the atomic level.

As seen in Fig. 2, the conduction process involves transitions between two main states with two (*a, c, e, f*) or three (*b, d*) K⁺ ions, respectively, in the selectivity filter. The ions proceed along the pore axis in a single-file fashion with a single water molecule between them, indicating that each permeating K⁺ ion is accompanied by about one water molecule in accord with streaming potential measurements¹⁵. The pathway with the lowest free energy barrier follows the sequence of configurations *a–b–d–f* (dotted line in the two-dimensional maps). Although all the microscopic elementary steps can occur reversibly, we describe and discuss the elementary steps leading to the outward movement of one ion for the sake of clarity. Along this optimal pathway, the ion in the cavity first approaches the intracellular entrance to the selectivity filter (*a–b*), then pushes the two ions in the selectivity filter (*b–d*) leading to the exit of the outermost ion on the extracellular side (*d–e–f*). These

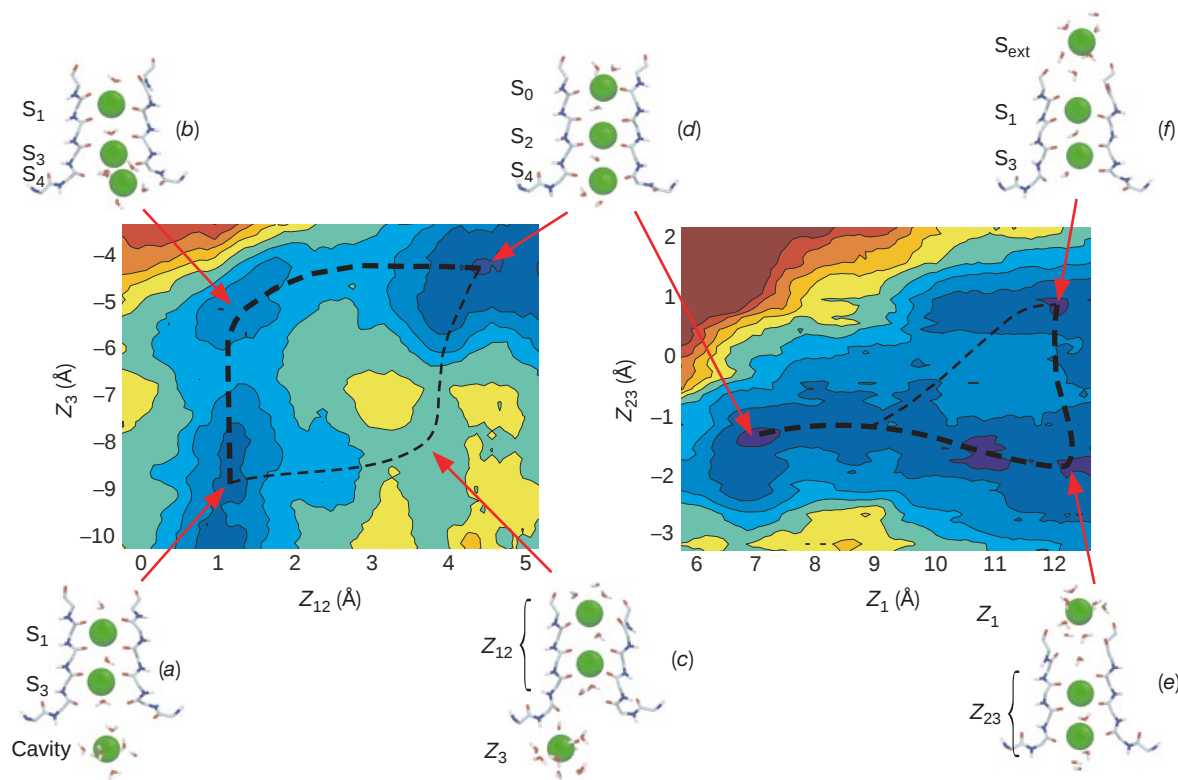


Figure 2 Topographic free energy maps of ion conduction calculated from umbrella sampling molecular dynamics simulations. Each colour level corresponds to an energy of 1 kcal mol^{−1}; the axes are in Å. The position of the ions, Z_1 , Z_2 and Z_3 (numbered in successive order starting from the outermost ion) are defined relative to the centre-of-mass of the backbone atoms of residues Thr 75–Val 76–Gly 77–Tyr 78, which constitute the KcsA central core of the selectivity filter. The definition of the reduced reaction coordinates (Z_{12} , Z_3) and (Z_1 , Z_{23}) is indicated in *c* and *e*, with Z_j corresponding to the

center of mass of ions i and j . The two-dimensional maps were calculated from a projection of the full three-dimensional PMF and a direct comparison of the energy levels in the two two-dimensional maps is thus meaningful. The lowest energy pathway (highlighted by a dotted line) follows the configurations *a–b–d–f* (a spontaneous transition along this pathway has been observed previously during an unbiased molecular dynamics simulation¹¹). A secondary pathway (highlighted by a thin dash line) following the configurations *a–c–d–f* is also possible.

elementary steps are reminiscent of the knock-on mechanism that was proposed nearly 50 years ago¹. The largest free energy barrier for this process is on the order of 2–3 kcal mol⁻¹, which is remarkable given the very large hydration energies of K⁺ ions. Once the outermost ion has exited on the extracellular side ($Z_1 \geq 12$ Å; Fig. 2, right two-dimensional map) the two ions remaining in the selectivity filter can move back and forth in the selectivity filter with a small barrier of approximately 1 kcal mol⁻¹ (e and f), recovering rapidly a productive configuration for conduction. The absence of any significant activation free energy barrier opposing the concerted ion translocation implies that ion conduction is essentially diffusion limited. A second pathway characteristic of a ‘vacancy-diffusion’ mechanism¹⁶ is also possible (sequence of configurations *a–c–d*, indicated with a thin, dashed line in Fig. 2). The two ions in the selectivity filter move first (*a–c*), leaving a vacant binding site formed by the carbonyl oxygens of Thr 75, which is subsequently occupied by the incoming ion from the cavity (*c–d*). The largest energy barriers for this process are on the order of 3–4 kcal mol⁻¹. The existence of multiple pathways may contribute to yielding a high throughput as ion translocation does not have to proceed in a unique fashion.

Even though no large free energy barrier opposing ion conduction is observed in the free energy maps, some positions along the permeation pathway are preferably occupied by K⁺ ions. These ‘binding sites’ are illustrated in Fig. 3 by showing a superposition of a few dynamical configurations associated with the free energy minima in the two dimensional maps. Five specific sites in the selectivity filter are revealed by the superposition. As expected, the sites (S₁–S₄), which were detected in the X-ray structure of KcsA, are observed⁴. Moreover, the calculations reveal the presence of two additional K⁺-binding sites, located on the extracellular side of the

channel, S₀ and S_{ext}. The K⁺ ion in site S₀ is hydrated by 3–4 water molecules and makes some contacts with the carbonyl of Tyr 78, whereas the ion in the outermost site, S_{ext}, is almost fully hydrated. These two binding sites were not detected in the crystallographic structure of KcsA⁴, but are now observed in recent diffraction data at higher resolution (R. MacKinnon and J. Cabral, personal communication). Sites S₀, S₂ and S₄ can be occupied simultaneously by K⁺ with a single water molecule located between each ion pair, for a total of three K⁺ ions in the selectivity filter (Fig. 2d). According to the free energy map shown in Fig. 2, such a configuration K-W-K-W-K corresponds to a shallow free energy well and does not oppose rapid conduction. Recent calculations have indicated that tetraethylammonium (TEA), a well known external blocker of K⁺ channels¹⁷, binds favourably in site S₀ in a K-W-K-W-TEA configuration¹⁸.

The details of the process of ion conduction challenge a naive view of repulsive forces in multi-ion transport systems. A particularly notable example is provided by configuration *b* in Fig. 2, in which one ion occupies the inner site S₄ while the other ion occupies the adjacent site S₃. The two ions, which are only 4 Å away from each other, appear to be stabilized by interacting simultaneously with the carbonyl groups of Thr 75. At such a distance the electrostatic energy is +83 kcal mol⁻¹, and the mutual repulsive coulombic force between the two K⁺ ions corresponds to an electric field of 9×10^9 V m⁻¹, which is equivalent to the field produced by a transmembrane potential of 27 V. However, the free energy of this configuration is only approximately 1 kcal mol⁻¹ higher than the reference configuration *a* according to the two-dimensional map in Fig. 2. Nonetheless, further analysis reveals that repulsive forces are absolutely essential for rapid conduction. Figure 4 shows the free energy profile as a function of the position of the outermost ion

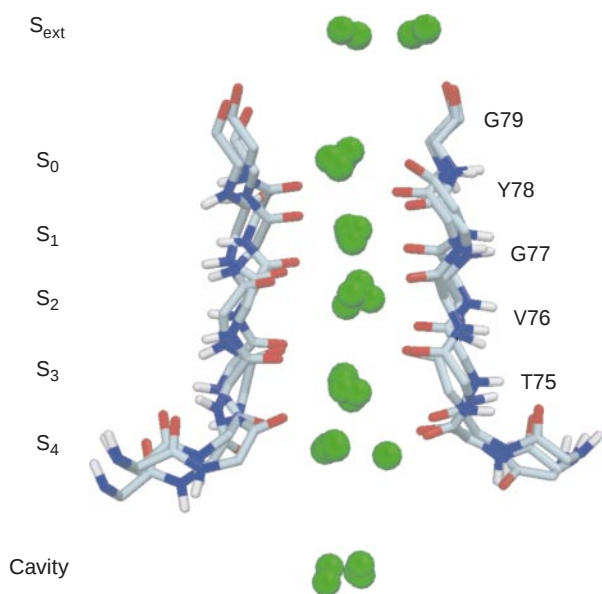


Figure 3 Superposition of instantaneous dynamical configurations showing the dominant ion positions associated with the free energy minima in Fig. 2. The outer site (S₁), the upper inner site (S₃), the lower inner site (S₄) and the diffuse cavity site detected in the initial X-ray structure are shown⁴. Furthermore, K⁺ can occupy the site where a crystallographic water was previously detected (S₂), as well as two additional sites termed S₀ and S_{ext}. The ions in S₁, S₂ and S₃ are in contact with the main chain carbonyl oxygens (S₁, Gly 77 and Tyr 78; S₂, Val 76 and Gly 77; S₃, Thr 75 and Val 76) and only 1–2 water molecules, whereas the ion in S₄ is hydrated by 2–3 water molecules and makes intermittent contacts with the carbonyl oxygens and the side chain of Thr 75. The ion in S₀ is in contact with Tyr 78 and is hydrated by 3–4 water molecules. The ion in S_{ext} is almost fully hydrated.

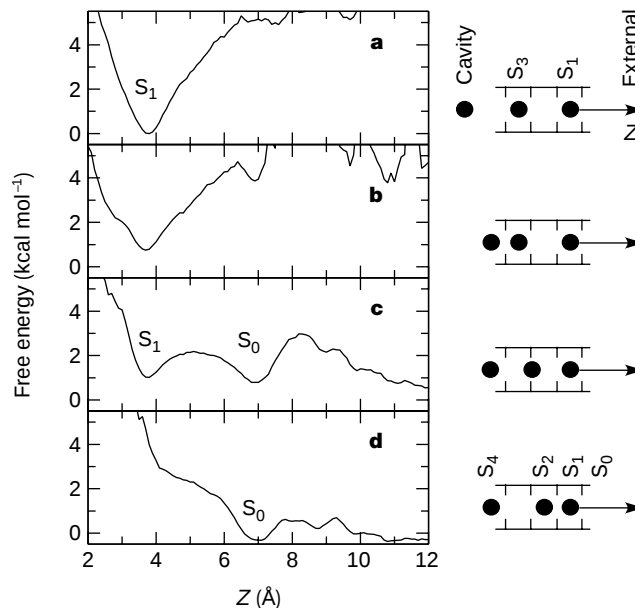


Figure 4 Importance of ion–ion repulsion on the elementary steps of ion conduction. The total free energy of the three-ion system is plotted as a function of the outermost K⁺ ion while the two other K⁺ ions are fixed at different position in the pore. The corresponding ion configurations of the system are shown schematically on the right. The profiles were extracted from the full three-dimensional free energy $W(Z_1, Z_2, Z_3)$ and the absolute energy scale was preserved (a direct comparison of the profiles is meaningful). **a**, A large free energy barrier prevents the exit of the outer K⁺ ion towards the extracellular side as the two other ions are in the cavity and in the inner site (S₃). **b, c**, The barrier towards the extracellular side is progressively reduced and site S₀ is slightly stabilized as the cavity ion approaches site S₄ and the other ion initiates a transition from site S₃ to S₂. **d**, Once two ions occupy sites S₄ and S₂, the energy well at site S₁ disappears and the outer ion transits to site S₀.

while the two other ions are fixed at different locations in the pore. Initially, while the two other ions are in the cavity and in the inner site S_3 , the outer K^+ ion is located in a deep well and cannot exit towards the extracellular side (Fig. 4a). As the ion in the cavity approaches the inner site S_4 and the second ion initiates a transition from site S_3 to S_2 , the bottom of the free energy well is lifted up, thereby decreasing the barrier between sites S_1 and S_0 progressively to 1 kcal mol⁻¹ (Fig. 4b, c). Ultimately, the site S_1 becomes unstable and the exit of the ion towards the extracellular side becomes barrierless (Fig. 4d). This series of free energy profiles demonstrates that without the repulsion from the other two incoming ions, the outer ion would be trapped in the site S_1 and its exit towards the extracellular side would require a significant activation free energy.

The calculations show that the conduction of K^+ ions is essentially barrierless. We next sought to determine whether the channel remains selective for K^+ ion. Selectivity arises primarily from differences between the free energy of the ion in the channel and in the bulk solution¹⁹. To address questions about selectivity, we performed molecular dynamics free energy simulations (MD/FES) in which one K^+ is alchemically transformed into a Na^+ at specific locations along the pore axis²⁰ (see Methods for details). The calculated excess free energy difference relative to the aqueous solution for the outer site (S_1) near the extracellular surface is about +2.8 kcal mol⁻¹. Moving deeper into the pore, the calculated free energy of the subsequent binding site (S_2) rises to +6.6 kcal mol⁻¹, indicating that there is an increase in the selectivity of the channel. The relative selectivity of the 'external lock-in site' detected in Ba^{2+} blockade experiments has been estimated to be around +5.5 kcal mol⁻¹ in one particular mammalian K^+ channel³. The calculations show that KcsA has both the ability to rapidly conduct K^+ ions and to discriminate against Na^+ . It should be emphasized that the magnitude of the dynamical fluctuations of the carbonyl oxygen atoms that form the selectivity filter (visible in the superposition of configurations shown in Fig. 3) is much larger than the difference in the radius of Na^+ and K^+ . This shows that selectivity does not arise from simple geometric considerations based on a rigid pore.

Further computational experiments indicate that the fluctuations of the protein structure can have a considerable influence on the free energy profiles and on the selectivity of the channel. Simulations with an artificial restraint distance of 4 Å imposed between the Tyr78 side chain (from the GYG signature sequence) and its hydrogen-bonding neighbour Trp68 exhibit a reduction in the fluctuations of the channel, giving rise to a destabilization of the two-ion state relative to the intermediate three-ion state by 2–3 kcal mol⁻¹. The relative stability of the two states is affected because the entropic contribution from the protein to the free energy profile (artificially reduced by the restraints) is larger when two ions are located in the selectivity filter. Moreover, the reduction in fluctuations also modifies the selectivity of the channel by increasing the excess free energy of Na^+ over K^+ in the selectivity filter. Thus, important properties of the pore are coupled to the fluctuations of the hydrogen bond between Tyr78 and Trp68, nearly 12 Å away. This may explain how even relatively conservative mutations at positions distant from the pore can markedly affect the conductance of K^+ channels²¹. It is possible that similar, long-range coupling mechanisms in which the free energy surface governing the translocation of ions is modified in response to a perturbation at a large distance (for example, the binding of ATP), may have an important role in the function of ion pumps²². □

Methods

Simulations and parameters

All simulations were carried out using the program CHARMM²³. The total number of atoms in the simulation system is slightly above 40,000 (KcsA, 112 dipalmitoyl phosphatidylcholine (DPPC), 6,532 water molecules, 3 K^+ in the pore, and 12 K^+ and 23 Cl^- in the bulk solution). The side chain of Glu 71 was constructed in a protonated state

forming a hydrogen bond with Asp 80. The channel axis is oriented along the Z axis; the centre of the membrane is at $Z = 0$. The simulation methodology has been described previously¹¹. Briefly, the electrostatic interactions were computed with no truncation using the particle mesh Ewald (PME) algorithm²⁴. The trajectories were generated with a time step of 0.2 fs at constant pressure (1 atm) and temperature (330 K)²⁵. We used the PARAM22 potential function for proteins²⁶, lipids²⁷ and the TIP3 water potential²⁸.

The potential function of the ions was calibrated to yield an accurate description of solvation in bulk water and throughout the selectivity filter of the K^+ channel, which is lined by backbone oxygens. The Lennard–Jones (LJ) parameters of the K^+ and Na^+ ions were adjusted to yield the experimental free energy in liquid water (that is, –80 and –101 kcal mol⁻¹), respectively. In addition, the LJ parameters for the cation–carbonyl oxygen pairs were refined to yield solvation free energies in liquid *N*-methylacetamide identical to those in bulk water. The parameters were refined using the spherical solvent boundary potential, which incorporates the long-range electrostatic reaction field²⁹.

Umbrella sampling PMF and selectivity calculations

For the umbrella sampling PMF calculations, a total of 308 independent simulations of 100 ps with a biasing harmonic potential centred first on Z_{12} and Z_3 (varying successively from 0 to 5.0 and –10.0 to –3.5 every 0.5 Å), and then on Z_1 and Z_{23} (varying successively from 6.0 to 12.5 and –3.0 to 2.0 every 0.5 Å) were generated with a force constant of 20 kcal mol⁻¹ Å⁻². The centre-of-mass biasing potentials were implemented using the MMFP option of CHARMM²³. The entire simulation time included in the PMF calculations is nearly 38 ns. The umbrella sampling simulations were unbiased together using the weighted histogram analysis method (WHAM)³⁰ to calculate the full PMF $W(Z_1, Z_2, Z_3)$. This was then projected onto the reduced reaction coordinates (Z_{12}, Z_3) and (Z_1, Z_{23}) to yield two-dimensional, free energy maps. The projection maintains the relative energy levels shown in the two two-dimensional maps.

For the selectivity calculations, each alchemical free energy difference between K^+ and Na^+ was obtained from forward and backward MD/FES trajectories, generated with the thermodynamic coupling parameter method²⁰, for a total of 440 ps. The total free energy difference was then calculated using WHAM³⁰. Similar calculations have been performed previously using reduced models of the KcsA channel^{9,10}.

Received 16 August; accepted 8 September 2001.

- Hodgkin, A. L. & Keynes, R. D. The potassium permeability of a giant nerve fibre. *J. Physiol.* **128**, 61–88 (1955).
- Hille, B. & Schwarz, W. Potassium channels as multi-ion single-file pores. *J. Gen. Physiol.* **72**, 409–442 (1978).
- Neyton, J. & Miller, C. Discrete Ba^{2+} block as a probe of ion occupancy and pore structure in the high-conductance Ca^{2+} -activated K^+ channel. *J. Gen. Physiol.* **92**, 569–586 (1988).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K^+ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
- Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–639 (1990).
- Guidoni, L., Torre, V. & Carloni, P. Potassium and sodium binding to the outer mouth of the K^+ channel. *Biochem.* **38**, 8599–8604 (1999).
- Shrivastava, I. H. & Sansom, M. S. Simulations of ion permeation through a potassium channel: molecular dynamics of KcsA in a phospholipid bilayer. *Biophys. J.* **78**, 557–570 (2000).
- Allen, T. W., Bliznyuk, A., Rendell, A. P., Kuyucak, S. & Chung, S. H. The potassium channel: structure, selectivity and diffusion. *J. Chem. Phys.* **112**, 8191–8204 (2000).
- Aqvist, J. & Luzhkov, V. Ion permeation mechanism of the potassium channel. *Nature* **404**, 881–884 (2000).
- Bernèche, S. & Roux, B. Molecular dynamics of the KcsA K^+ channel in a bilayer membrane. *Biophys. J.* **78**, 2900–2917 (2000).
- Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300–313 (1935).
- Roux, B. Statistical mechanical equilibrium theory of selective ion channels. *Biophys. J.* **77**, 139–153 (1999).
- Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comp. Phys.* **23**, 187–199 (1977).
- Alcayaga, C., Cecchi, X., Alvarez, O. & Latorre, R. Streaming potential measurements in Ca^{2+} -activated K^+ channels from skeletal and smooth muscle. Coupling of ion and water fluxes. *Biophys. J.* **55**, 367–371 (1989).
- Schumaker, M. F. & MacKinnon, R. A simple model for multi-ion permeation. single-vacancy conduction in a simple pore model. *Biophys. J.* **58**, 975–984 (1990).
- Armstrong, C. M. Induced inactivation of the potassium permeability of squid axon membranes. *Nature* **219**, 1262–1263 (1968).
- Crouzy, S., Bernèche, S. & Roux, B. Extracellular blockade of K^+ channels by TEA: results from molecular dynamics simulations of the KcsA channel. *J. Gen. Physiol.* **118**, 207–218 (2001).
- Eisenman, G. Cation selective electrodes and their mode of operation. *Biophys. J.* **2** (Suppl. 2), 259–323 (1962).
- Kollman, P. A. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).
- Perozo, E., MacKinnon, R., Bezanilla, F. & Stefani, E. Gating currents from a nonconducting mutant reveal open-closed conformations in Shaker K^+ channels. *Neuron* **11**, 353–358 (1993).
- Lauger, P. A channel mechanism for electrogenic ion pumps. *Biochim. Biophys. Acta* **552**, 143–161 (1979).
- Brooks, B. R. *et al.* CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
- Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
- Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation—the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

26. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
27. Schlensker, M. J., Brickmann, J., MacKerell, A. D. Jr & Karplus, M. in *Biological Membranes. A Molecular Perspective from Computation and Experiment* (eds Merz, K. M. & Roux, B.) 31–81 (Birkhauser, Boston, 1996).
28. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
29. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063 (1994).
30. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.* **13**, 1011–1021 (1992).

Acknowledgements

Discussions with R. MacKinnon and J. Cabral are gratefully acknowledged. This work was supported by the National Institutes of Health and by the Canadian Institutes of Health Research. Several calculations were performed at the RQCHP Computer Center of the University of Montreal and at the National Center for Supercomputing Applications (NCSA) of the University of Illinois at Urbana-Champaign. We are grateful to B. Lorazo for his support. All the molecular pictures were drawn with DINO (A. Philippsen).

Correspondence and requests for materials should be addressed to B.R. (e-mail: benoit.roux@med.cornell.edu).

Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion

C. Erec Stebbins* & Jorge E. Galán

Section of Microbial Pathogenesis, Boyer Center for Molecular Medicine, Yale School of Medicine, New Haven, Connecticut 06536, USA

Many bacterial pathogens use a type III protein secretion system to deliver virulence effector proteins directly into the host cell cytosol, where they modulate cellular processes^{1,2}. A requirement for the effective translocation of several such effector proteins is the binding of specific cytosolic chaperones, which typically interact with discrete domains in the virulence factors^{3,4,5}. We report here the crystal structure at 1.9 Å resolution of the chaperone-binding domain of the *Salmonella* effector protein SptP with its cognate chaperone SicP. The structure reveals that this domain is maintained in an extended, unfolded conformation that is wound around three successive chaperone molecules. Short segments from two different SptP molecules are juxtaposed by the chaperones, where they dimerize across a hydrophobic interface. These results imply that the chaperones associated with the type III secretion system maintain their substrates in a secretion-competent state that is capable of engaging the secretion machinery to travel through the type III apparatus in an unfolded or partially folded manner.

Comprising more than 20 proteins and closely related to the flagellar assembly apparatus, type III systems stand among the most complex protein secretion systems known^{1,2}. A distinguishing essential feature of these systems is the requirement of a family of customized cytoplasmic chaperones for the secretion of their cognate substrate effector proteins^{3,4,5}. Each chaperone is specific in most cases for one or, in a few cases, two secreted proteins. Therefore, absence of a given chaperone affects the secretion only of its cognate secreted protein, which prematurely degrades or accumulates within the bacterial cytoplasm. Although there is little

amino-acid sequence similarity among these chaperones, they share several features such as a comparatively small relative molecular mass (<20,000; M_r <20K), a generally low isoelectric point and a secondary structure predicted to be predominantly helical. The actual function of these chaperones is poorly understood. SptP is a tyrosine phosphatase⁶ and GTPase activating protein (GAP) for Cdc42 and Rac^{7,8} that reverses the changes induced by *Salmonella* on entry into host cells⁹. SicP is an acidic protein of M_r 14K required for the stability of SptP within the bacterium and its secretion through the type III secretion system¹⁰. Previous deletion analysis established that SptP residues 15–100 were sufficient to bind SicP¹⁰, but protease footprinting indicates that a different polypeptide segment is protected by the chaperone (Fig. 1a and b). The structure of this protected complex was determined by multiple-wavelength anomalous dispersion (MAD) from the scattering of 31 bromine ions

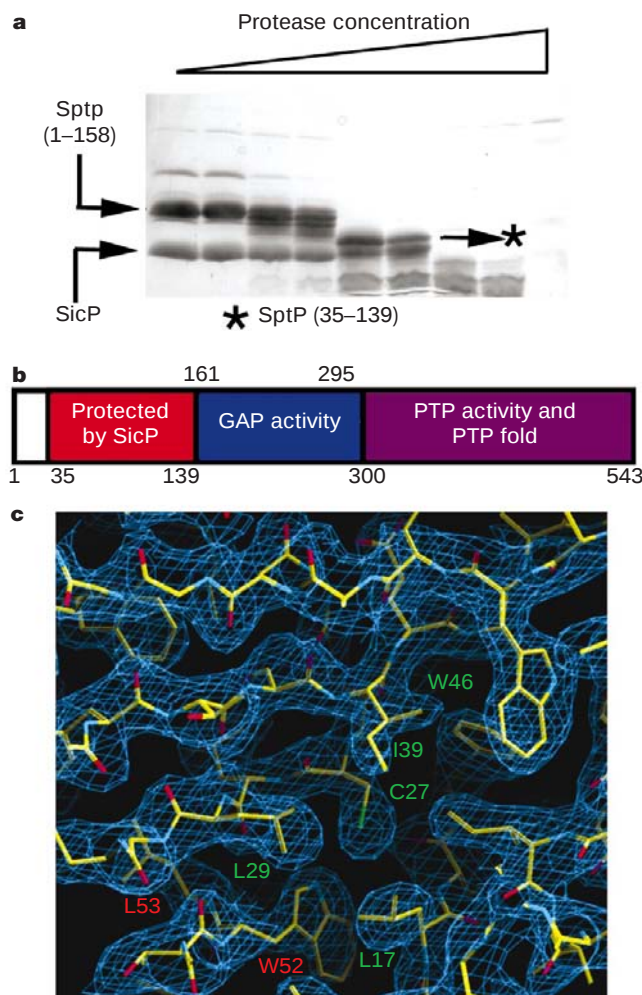


Figure 1 Protease footprinting delineates the chaperone-binding sequence of SptP. **a**, Purified SicP–SptP^{1–158} was incubated on ice for 30 min with increasing concentrations of the protease subtilisin. A fragment of SptP (asterisk) was protected by SicP, and determined by N-terminal sequencing and mass spectroscopy to span SptP residues 35–139. **b**, Domain schematic of SptP. The protected, chaperone-binding domain is shown in red, and the enzymatic effector domains in blue (for the GAP) and purple (for the tyrosine phosphatase domain). The amino-acid numbers that mark the beginning and end of each domain are labelled. **c**, Electron density from a map calculated with MAD solvent-flattened phases to 2.2 Å resolution and contoured at 1.0σ. Residues from the final, refined models of SptP and SicP are shown with atoms of oxygen, nitrogen and carbon in red, blue and yellow, respectively. Some amino-acid side chains around W52 of SptP are labelled in green (SicP) or red (SptP).

* Present address: Laboratory of Structural Microbiology, The Rockefeller University, Box 52, 1230 York Avenue, New York, New York 10021, USA.

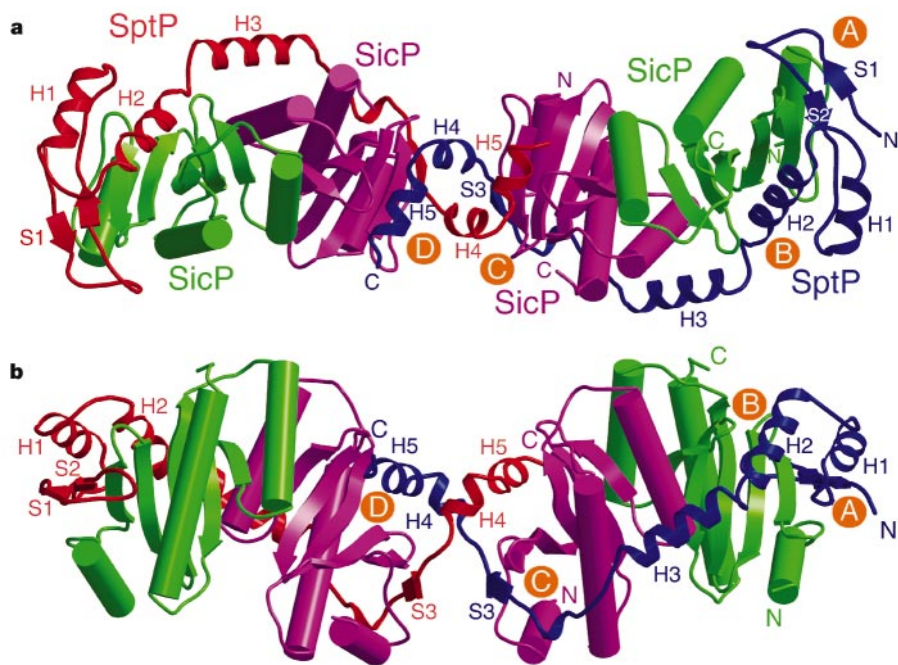


Figure 2 Each molecule of SptP binds as an unfolded polypeptide to three SicP chaperones. Two views (**a** and **b**) of the SptP–SicP complex as cartoons related by a 90° rotation about a horizontal axis. SicP is shown with α -helices as solid cylinders, and β -strands as thick arrows. The helices of SptP are shown as ribbons, with the two different polypeptide chains shown in red and blue. The non-interacting middle two molecules of SicP at the centre of the complex are shown in magenta, and the outer two SicP molecules of the tetrameric chaperone arrangement are shown in green. The SicP chaperone

homodimer pairs are therefore mixed with one green and one magenta molecule. The secondary structural elements of SptP are labelled, and the four important regions of contact between SptP and SicP are labelled A–D in orange (for one SptP molecule). Regions C and D at the carboxy terminus of the chaperone-binding domain are separated from regions A and B by a long α -helix (H3) that makes few contacts with SicP, and runs across the homodimer pair from the outer SicP chaperone to the middle chaperone of one homodimer pair.

introduced into the crystal by short duration soaks in concentrated halide solution (Fig. 1c and Table 1). The structure of the SptP–SicP complex is remarkable in many ways. The asymmetric unit of the crystals contains four molecules of the chaperone SicP, aligned in a linear fashion and arranged in two sets of tightly bound homodimers that bind two SptP molecules (Fig. 2). The SicP homodimer pairs do not interact with each other, but are held together by a molecular interface formed between the two SptP molecules. The overall complex is characterized by nearly perfect two-fold symmetry about an axis that runs through the centre of the SptP dimerization interface. The chaperone-binding domain of SptP does not adopt a globular fold for interaction with SicP. Instead, each SptP molecule

is seen in the crystals as a non-globular polypeptide that retains secondary structure and is wrapped around three SicP chaperones (Fig. 2). Because the chaperone-binding domain is not globular, the interaction of a single molecule of SptP with three SicP chaperones results in a large (6,200 Å²) surface area being buried on formation of this complex. Numerous specific hydrophobic, polar and salt-bridged interactions maintain the chaperone-binding domain tightly bound to the surface of the SicP molecules. Beginning from the distal ends of the four-chaperone linear arrangement, each chaperone-binding domain of SptP snakes its way around opposite sides of the different chaperone homodimer pairs, approaching each other at the centre of the complex, where they interdigitate and form an intimate molecular interface composed

Table 1 Summary of crystallographic analysis

Diffraction data								
Crystal	Wavelength (Å)	Resolution (Å)	Unique Reflections	Measured (%)	Coverage	Data statistics		
						R_{sym}		$I/\sigma(I)$
NaBr (peak)	0.9212	2.50	33,933	511,486	100.0 (100.0)	7.1 (21.5)		29.8 (15.2)
NaBr (inflection)	0.9215	2.50	33,562	254,600	100.0 (100.0)	6.1 (18.2)		20.6 (8.1)
NaBr (remote)	0.9050	2.50	33,043	238,218	99.9 (100.0)	5.7 (15.9)		22.1 (8.3)
Native	1.0051	1.90	75,880	452,315	96.2 (74.3)	4.5 (24.3)		21.5 (3.7)
Mean figure of merit for MAD phasing: 0.35 (0.18)								
Refinement and analysis of molecular models								
Crystal	Resolution (Å)	Reflections ($ F \geq 0$)	Atoms modelled (protein/water)	R/R_{free} (%)	Bond length (Å)	Root mean square deviation		B-factor (Å ²)
						Angle (°)		
Native	30.0–1.9	70,014	5,712/629	22.4/25.8	0.006	1.2		1.4

$R_{\text{sym}} = \Sigma h \Sigma i |I_{h,i} - I_h| / \Sigma h \Sigma i I_{h,i}$, for the intensity (I) of i observations of reflection h . Values in parentheses are for the high-resolution shell. $R = \Sigma |F_o - F_{\text{calc}}| / \Sigma F_o$; F_{calc} is the model structure factor and 5% of the data are omitted for R_{free} . Bond and angle deviations are from ideal values; B-factor deviations are between bonded atoms.

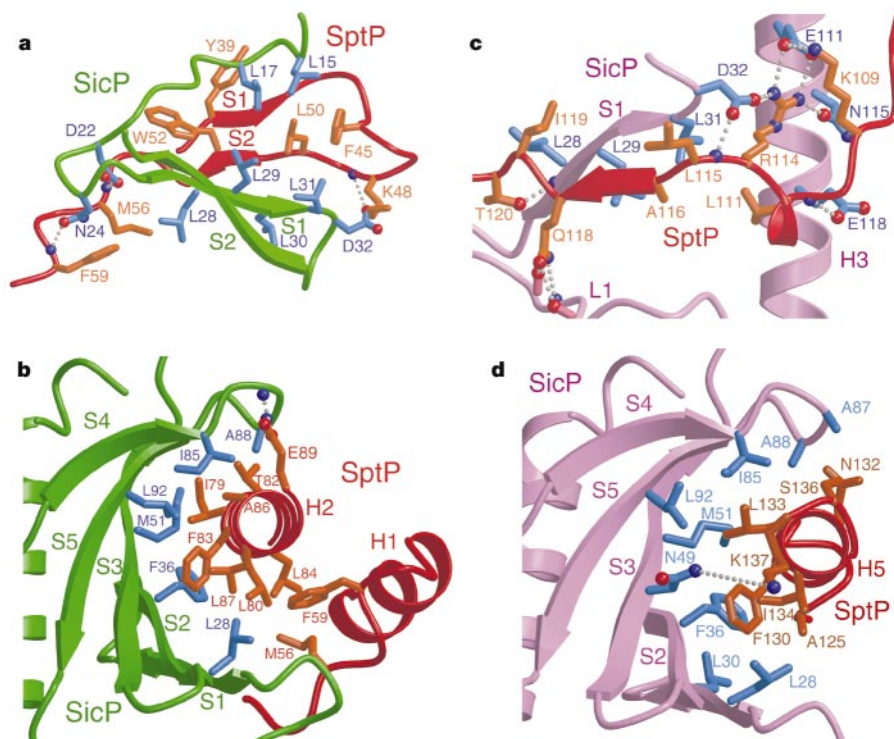


Figure 3 Different portions of SptP bind identical motifs in SicP primarily through hydrophobic interactions. **a**, An N-terminal portion of SptP (red with orange side chains, residues 39–53) forms an intermolecular antiparallel β -sheet (SptP strands S1 and S2) with an outer SicP molecule (green with blue side chains). Large hydrophobic residues of SptP (F45, W52 and F59) insert deeply into the hydrophobic surface patch of SicP. Atoms of oxygen and nitrogen are shown in red and blue, respectively, with hydrogen bonds as grey dotted lines. **b**, An α -helix of SptP (red with orange side chains) inserts into a hydrophobic groove (helix-binding groove) formed from the highly twisted β -sheet of an outer SicP molecule (green with blue side chains). The helix-binding groove surrounds the helix, thereby contacting almost all SptP side chains along the length of the helix. SptP contacts are primarily hydrophobic, and are centred on a phenylalanine stacking between

residues Phe 83 of SptP and Phe 36 of SicP. **c**, A C-terminal portion of SptP (red with orange side chains) binds as an extended structure and forms an intermolecular antiparallel β -sheet with a middle SicP molecule (magenta with blue side chains). Before its strand pairing sequence, SptP anchors this region of structure to SicP through the deep insertion of Arg 114, which hydrogen bonds with residues Asp 32 and Glu 111 of SicP. **d**, Having crossed over from one SicP homodimer pair to the other, a C-terminal α -helix of SptP (red with orange side chains) inserts into the helix-binding groove of a middle SicP molecule (magenta with blue side chains). The interaction of this helix with the SicP groove, like the one between the N-terminal H2 helix of SptP with an outer SicP molecule, is centred on the stacking of Phe 36 of SicP with a phenylalanine of SptP (Phe 130 for the C-terminal helix of the SptP chaperone-binding domain).

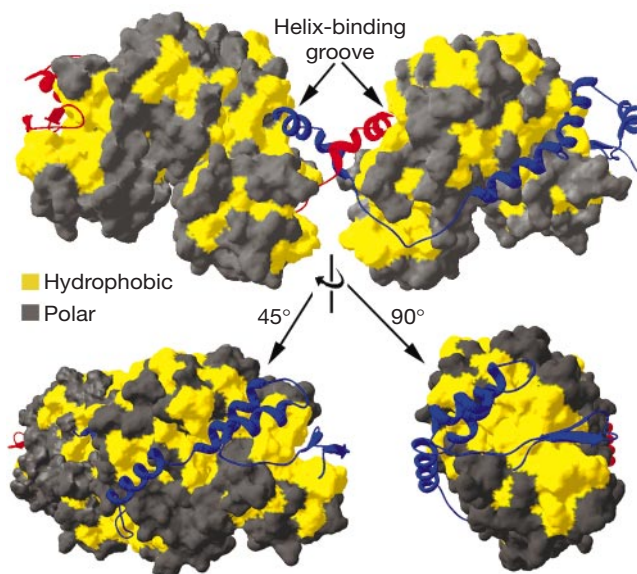


Figure 4 The tetrameric SicP arrangement presents an extensively hydrophobic surface that is buried on SptP binding. Three views of the complex are shown, with SicP shown as a molecular surface with polar regions grey and hydrophobic regions yellow. SptP is shown as a ribbon diagram, as in Fig. 2. The top view is oriented identically to Fig. 2a, and

the two lower views are related by rotations about a vertical axis by 45° and 90°. For clarity, the many side chains of SptP that cover the hydrophobic surface of SicP are not shown.

mostly of hydrophobic interactions. After crossing each other, the remaining portions of the SptP molecules interact with a third chaperone in the opposite homodimeric pair (Fig. 2). This interaction occurs through an α -helix in SptP, which fits in a pronounced groove formed by a highly twisted β -sheet in SicP (henceforth, helix-binding groove). The same helix-binding groove is used by the outermost chaperones of the tetrameric chaperone complex to bind a different helix located at the amino terminus of SptP. Although the overall features of these helices are similar and both exhibit a phenylalanine that is deeply buried in the same place in the SicP groove, they do not share significant sequence identity. The same theme of identical regions of SicP binding different regions of SptP occurs again in the top two strands of the helix-binding groove, which in the central chaperones bind a region at the carboxy terminus of the chaperone-binding domain of SptP and in the outer chaperones interact with its N terminus (Fig. 2).

SptP interacts with the SicP chaperone dimers mainly through four regions of its chaperone-binding domain (labelled A–D in Fig. 2 and shown in detail in Fig. 3). In region A, two antiparallel β -strands lie on top of SicP, forming an intermolecular β -sheet (Fig. 3a). Numerous hydrophobic and polar interactions between SptP and SicP side chains cement the proteins together in this region (Fig. 3a). Region B is centred on the insertion of the H2 helix of SptP into the helix-binding groove of an outer molecule of SicP, and a stacking of phenylalanines from both molecules (Fig. 3b). The groove wraps around the SptP H2 helix so extensively that nearly every side chain in this helix is able to contact residues in SicP. Region C interacts only with the middle chaperone across an antiparallel intermolecular strand pairing of SptP S3 to SicP S1 (Fig. 3c). The same strand of SicP (S2) in the outer chaperone molecule is used to strand pair with an N-terminal region of the chaperone-binding domain (the SptP S2 strand). Dominating the region C interaction is an extended series of hydrogen bond contacts centred on the deep insertion of Arg 114 of SptP into a binding cleft formed by SicP (Fig. 3c). Finally, region D is situated on the opposing middle chaperone (Fig. 2). Following region C, SptP crosses over to region D through its dimerization sequence, and inserts a C-terminal helix into the helix-binding groove of the opposing chaperone. The H5 helix of SptP is almost completely surrounded by the groove, and is also centred on the stacking of phenylalanines in the helix-binding groove (Fig. 3d). Following a series of hydrophobic and polar contacts over the remainder of the helix, SptP exits the groove and its C terminus lies just out of contact with the chaperone.

The SptP-interacting surfaces of the chaperone are predominantly hydrophobic (Fig. 4). The overall architecture of the SptP–SicP complex suggests a stepwise engagement of the unfolded polypeptide by the chaperone. In this model, the assembly is nucleated on the engagement of the four key hydrophobic regions of SptP by the binding crevices of SicP. The size and hydrophobic nature of the residues that dominate these regions of SptP confer a high binding energy on complex formation. These four contact points serve as anchors to latch SptP to SicP, and allow the remainder of the contacts to form, zipping the SptP polypeptide over the chaperone molecules.

In the crystal structure, there are two different homodimeric interfaces, one between the C-terminal regions of SptP and the other between SicP (Fig. 5). The dimerization interface of SptP, involving residues 119–135 (and burying 900 Å² of surface area), is located in the centre of the complex and is predominantly hydrophobic. Because the SicP–SptP complexes are observed by gel filtration chromatography in either a 2:1 or 4:2 ratio (data not shown), the significance of this interface is unclear. The presence of an intermolecular disulphide bond linking Cys 122 of each molecule at the centre of the SptP dimer interface (not observed in any complexes before crystallization; data not shown) raises the possibility that this smaller interface may occur only in the crystals, or, in general, at

high protein concentration. It is therefore possible that the H5 helices, which are ‘crossed over’ to bind the third SicP molecule in the 4:2 complex, may bind the same helix-binding groove but in the opposite, symmetrically placed SicP to form a 2:1 complex. It is unclear what conformation the dimerization domain of SptP would adopt in this model, but the polypeptide would require flipping its directionality as the helix-binding grooves of the two middle chaperones are oppositely oriented. In this scheme, the core functional unit would be a SicP homodimer binding one unfolded SptP

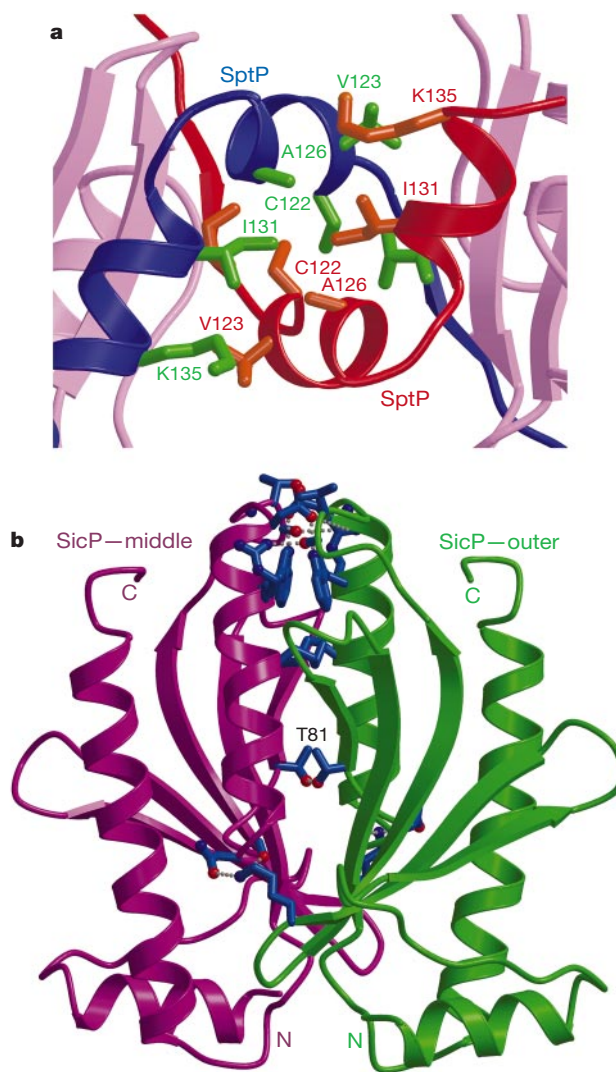


Figure 5 The hydrophobic dimerization of SptP and the extensive homodimeric SicP interface hold the complex together. **a**, The SptP dimerization interface, located at the centre of the complex on the two-fold axis of symmetry, is formed as the C-terminal regions of SptP cross from one SicP homodimer pair to another. Between the two middle chaperones, the last two SptP helices interact primarily through hydrophobic contacts. Side chains from the helices that interact with SicP are omitted for clarity. The two molecules of SptP (with orange and green side chains) are again shown in red and blue, and the surrounding portions of the middle chaperones are shown in magenta. The N and C termini of each molecule are indicated. **b**, The homodimeric interface between a middle (magenta) and outer (green) SicP chaperone is extensive, harbouring points of contact from the ‘top’ to the ‘bottom’ of the molecule along the two-fold axis of symmetry (which is distinct from the global two-fold present in the complex). At the centre of the interface is a strong hydrogen bond between the Thr 81 side chains of each molecule. A large cluster of contacts also occurs at opposite ends of the interface, involving a large subset of residues, many of which form intermolecular hydrogen bonds.

chaperone-binding domain. It is also possible that both the 4:2 and 2:1 complexes are physiologically relevant and serve different functions during the secretion process.

The interface for the SicP homodimer is extensive, burying more than 2,100 Å² of surface area, and involves the bringing together of identical surfaces in each monomer through a two-fold axis of symmetry (Fig. 5). It consists of a mixture of hydrophobic contacts and hydrogen bonds, although this interface is significantly more hydrophilic than that between the chaperone and the chaperone-binding domain. Contacts include a large network of van der Waals interactions and hydrogen bonds and a strong (2.4 Å) hydrogen bond in the middle of the interface linking Thr81 from each molecule.

Recently, the crystal structure of the chaperone-binding domain of the type III secreted protein YopH was solved^{11,12}. Unlike the corresponding domain of SptP, the chaperone-binding domain of YopH is soluble in the absence of cognate binding proteins, and alone it folds into a compact, globular structure. It is not known whether, when bound to its cognate chaperone SycH, this domain of YopH is partially or totally unfolded in a manner analogous to what we observed in the case of StpP.

The maintenance of the chaperone-binding domain of SptP in an unfolded conformation through its binding to the cognate chaperone SicP has profound implications not only for the function of chaperones associated with type III secretion, but also for the potential mechanisms by which the secreted proteins are engaged by the secretion machinery. Our studies favour the hypothesis that SicP, and by extension other functionally related chaperones, aids the secretion process by maintaining their cognate target proteins in a secretion-competent conformation. The secretion-competent conformation would be one of an unfolded polypeptide that yet maintains secondary structure when thoroughly engaged by its cognate chaperone. By extension, our findings support the idea that polypeptides travel through the secretion channel in an unfolded or partially folded manner. This hypothesis would be in keeping with the postulated size of the translocation channel of the needle complex, the diameter of which has been estimated to be ~3 nm^{13,14}, too narrow to fit most completely folded polypeptides.

The generality of an unfolded chaperone-binding domain in the secretion of type III effector proteins remains to be established. These studies suggest, however, that this highly conserved delivery system for bacterial virulence factors translocates polypeptides that are actively maintained in an unfolded state. Such results not only aid in the understanding of type III secretion but may also open avenues for pharmacological intervention into this virulence mechanism. □

Methods

Protein purification and domain delineation

An N-terminal glutathione-S-transferase (GST) fusion of the *Salmonella typhimurium* protein SicP was cloned into an engineered bicistronic expression plasmid¹⁵ along with SptP (residues 1–158); it was affinity purified with a glutathione-sepharose resin (Pharmacia), and separated from GST by thrombin cleavage. The complex was further purified by ion exchange chromatography and subjected to limited proteolysis with subtilisin. The chaperone was resistant to digestion, but SptP^{1–158} was quickly chased to a protease-resistant fragment identified by N-terminal sequencing and mass spectroscopy as SptP residues 35–139. SptP^{35–139} was cloned into the bicistronic plasmid with GST–SicP, and the resulting complex was purified as described above. The preparation was then concentrated by ultrafiltration for crystallization to 60 mg ml⁻¹ in a buffer containing 10 mM Tris at pH 8.0, 250 mM sodium chloride and 2 mM dithiothreitol (DTT).

Crystallization and structure determination

Crystals of the SptP–SicP complex were grown by vapour diffusion using hanging drops formed from mixing a 1:1 volume ratio of 60 mg ml⁻¹ complex with an equilibration buffer consisting of 2 M sodium chloride and 5–10% polyethylene glycol molecular mass 6000 (PEG6000) supplemented with 2 mM DTT and 15% glycerol. Crystals formed in the space group C2 with cell dimensions $a = 167.5$ Å, $b = 49.8$ Å and $c = 137.8$ Å, $\alpha = 90.0^\circ$, $\beta = 121.9^\circ$, and $\gamma = 90.0^\circ$. After the determination of the structure, the crystals were shown to contain a single complex in the asymmetric unit composed of SicP and SptP in a ratio of 4:2 (for a total M_r of 85K in the asymmetric unit). Crystals tolerated 30-s soaks in 2 M

sodium bromide and 5–10% PEG6000 supplemented with 15% glycerol, and data on such soaks were collected at Brookhaven National Synchrotron Light Source beamline X12C as a MAD experiment at energies at the inflection point, peak, and high-energy remote regions around the bromine edge. Data were processed with the HKL software package¹⁶. The automated heavy-atom search and refinement program Solve¹⁷ located 31 bromine sites for an overall figure of merit of 0.35–2.5 Å resolution. Maximum entropy solvent flattening with Resolve¹⁸ produced interpretable electron density maps of high quality. These phases were used with a data set collected at CHESS beamline F1, and extended to 1.9 Å resolution. A partial model of more than 500 amino acids (of a final 730) was built automatically with wARP¹⁹, and the remainder of the protein built over several cycles of model building with program O²⁰ and refinement with program CNS²¹. The final model possesses good stereochemistry (90% of the residues in the most favoured region of the Ramachandran plot), and R/R_{free} values of 22.4% and 25.8%, respectively. Molecular graphics were generated with the programs Molscript²², Raster3D²³ and Swiss-pdb Viewer²⁴.

Received 2 August; accepted 19 September 2001.

- Galán, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 322–328 (1999).
- Cornelis, G. R. & Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **54**, 735–774 (2000).
- Wattiau, P., Bernier, B., Deslée, P., Michiels, T. & Cornelis, G. R. Individual chaperones required for Yop secretion by *Yersinia*. *Proc. Natl Acad. Sci. USA* **91**, 10493–10497 (1994).
- Wattiau, P., Woestyn, S. & Cornelis, G. R. Customized secretion chaperones in pathogenic bacteria. *Mol. Microbiol.* **20**, 255–262 (1996).
- Bennett, J. C. Q. & Hughes, C. From flagellum assembly to virulence: the extended family of type III export chaperones. *Trends Microbiol.* **8**, 202–204 (2000).
- Kaniga, K., Uralil, J., Bliska, J. B. & Galán, J. E. A secreted tyrosine phosphatase with modular effector domains encoded by the bacterial pathogen *Salmonella typhimurium*. *Mol. Microbiol.* **21**, 633–641 (1996).
- Fu, Y. & Galán, J. E. A salmonella protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401**, 293–297 (1999).
- Stebbins, C. E. & Galán, J. E. Modulation of host signaling by a bacterial mimic. Structure of the *Salmonella* effector SptP bound to Rac1. *Mol. Cell* **6**, 1449–1460 (2000).
- Galán, J. E. & Zhou, D. Striking a balance: modulation of the actin cytoskeleton by *Salmonella*. *Proc. Natl Acad. Sci. USA* **97**, 8754–8761 (2000).
- Fu, Y. & Galán, J. E. Identification of a specific chaperone for SptP, a substrate of the centisome 63 type III secretion system of *Salmonella typhimurium*. *J. Bacteriol.* **180**, 3393–3399 (1998).
- Evdokimov, A. G., Tropea, J. E., Routzahn, K. M., Copeland, T. D. & Waugh, D. S. Structure of the N-terminal domain of *Yersinia pestis* YopH at 2.0 Å resolution. *Acta Crystallogr. D* **57**, 793–799 (2001).
- Smith, C. L., Khandelwal, P., Keliikuli, K., Zuiderweg, E. R. P. & Saper, M. A. Structure of the type III secretion and substrate-binding domain of *Yersinia* YopH phosphatase. *Mol. Microbiol.* (in the press).
- Kubori, T. et al. Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602–605 (1998).
- Blocker, A. et al. The tripartite type III secretion of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *J. Cell Biol.* **147**, 683–693 (1999).
- Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL–ElonginC–ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Terwilliger, T. C. Maximum likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
- Perrakis, A., Morris, R. M. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunker, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots or protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Murphy, M. E. Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
- Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).

Acknowledgements

We thank the following people for access to and assistance with crystallographic equipment: at Yale University School of Medicine, P. Pepin of the Macromolecular X-ray Crystallography Facility, and S. Kunchaparty and A.-M. Quinn of the Research Computing Group for workstation access; at the CHESS, D. Szebenyi and the MacChes staff; B. Sweet, A. Saxena and the staff of Brookhaven beamline X12C. We also acknowledge N. Papavasiliou, P. Jeffrey, S. Fugmann and members of the Galán laboratory for discussions and critical reading of this manuscript before submission. C.E.S. was supported by a fellowship of the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation. This work was funded by Public Health Services grants to J.E.G.

Correspondence and requests for materials should be addressed to J.E.G. (e-mail: jorge.galan@yale.edu). Coordinates have been deposited in the Protein Data Bank under accession number RCSB014340.

In vitro evolution suggests multiple origins for the hammerhead ribozyme

Kourosh Salehi-Ashtiani & Jack W. Szostak

Howard Hughes Medical Institute, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

The hammerhead ribozyme was originally discovered in a group of RNAs associated with plant viruses, and has subsequently been identified in the genome of the newt (*Notophthalmus viridescens*)^{1,2}, in schistosomes³ and in cave crickets (*Dolichopoda species*)⁴. The sporadic occurrence of this self-cleaving RNA motif in highly divergent organisms could be a consequence of the very early evolution of the hammerhead ribozyme^{4,5}, with all extant examples being descended from a single ancestral progenitor. Alternatively, the hammerhead ribozyme may have evolved independently many times. To better understand the observed distribution of hammerhead ribozymes, we used *in vitro* selection to search an unbiased sample of random sequences for comparably active self-cleaving motifs. Here we show that, under near-physiological conditions, the hammerhead ribozyme motif is the most common (and thus the simplest) RNA structure capable of self-cleavage at biologically observed rates. Our results suggest that the evolutionary process may have been channelled, in nature as in the laboratory, towards repeated selection of the simplest solution to a biochemical problem.

We designed a selection scheme for the isolation of self-cleaving RNAs with the goal of isolating ribozymes that are highly active at near-physiological pH and concentration of divalent ions. We started with a DNA pool coding for large random-sequence RNAs (Fig. 1) to allow for the isolation of complex self-cleaving molecules. Self-cleavage during transcription has limited the ability of previous *in vitro* selection experiments to evolve highly active self-cleaving ribozymes⁶. We inhibited unwanted premature self-cleavage by including in the transcription reaction a high concentration of a 16-base oligodeoxynucleotide that was complementary to the designed cleavage site. Control experiments showed that self-cleavage of a hammerhead ribozyme during transcription decreased from 90% to essentially undetectable levels in the presence of 50 μ M of the inhibitory oligonucleotide. We then purified, by denaturing gel electrophoresis, full-length RNA transcripts free of the inhibitory oligonucleotide. The self-cleavage reaction was subsequently initiated by the addition of magnesium ions, and terminated by the addition of EDTA and denaturant. Molecules that had self-cleaved within a 16-nucleotide region near the 5' end of the pool were then

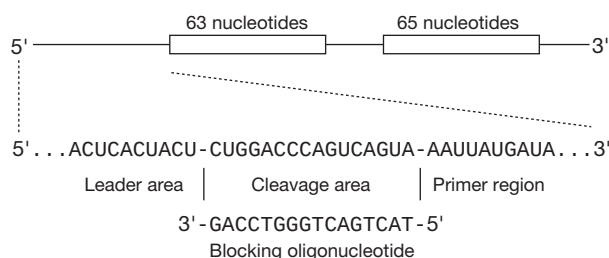


Figure 1 Pool design. Pool RNAs consisted of a 60-nucleotide 5' constant region (containing the cleavage region), two random regions (63 and 65 nucleotides long) interrupted by an 18-nucleotide constant region, and an 18-nucleotide 3' constant region. The sequences of the central portion of the 5' constant region and of the blocking oligonucleotide are shown.

selected on the basis of their altered electrophoretic mobility. The selected RNAs were reverse transcribed (RT), the complementary DNA was amplified by polymerase chain reaction (PCR), and the DNA transcribed to re-initiate the selection cycle.

After five rounds of selection, a low level of self-cleavage activity was observed in the pool, corresponding to an average rate of $\sim 0.003 \text{ min}^{-1}$ (Fig. 2a). In subsequent rounds, the reaction-incubation time was gradually reduced to enrich for more active molecules. By round 12, the pool activity had increased by almost 100-fold compared with round 5 (Fig. 2b). After round 12, pool molecules were mutagenized and subjected to additional rounds of selection. Four more rounds were carried out at lower pH and magnesium concentration (pH 7.2 and 0.5 mM MgCl_2); a further modest increase in activity to $\sim 0.8 \text{ min}^{-1}$ was observed by round 14 (Table 1).

We examined the evolution of the population in two ways. An initial analysis of the RNA cleavage sites showed that the ribozyme population changed significantly during the course of the selection (Fig. 2c). The round 5 RNAs cleaved largely in a cluster of sites in the 5' half of the 'cleavage area' (defined by the inhibitory oligonucleotide and the size of the gel-fractionated selected RNAs). However, as the selection progressed, the cleavage sites shifted mainly to a single site in the 3' region of the cleavage area in rounds 11 and 12, and then migrated again to the most 5' region of the cleavage area in rounds 14 and 16. To obtain a more detailed picture of the evolving RNA population, we cloned and sequenced 30–50 RNAs from rounds 5, 7, 9, 11, 12, 14 and 16. In round 5, when ribozyme

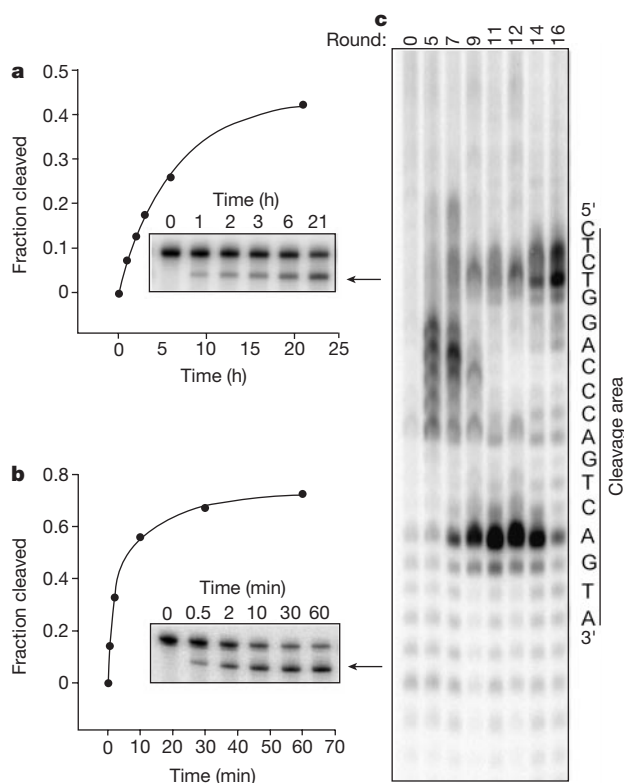


Figure 2 Self-cleavage of pool RNAs. After completion of each round, RNAs were transcribed in the presence of α - ^{32}P -UTP and blocking oligonucleotide, then purified and assayed for self-cleavage in 5 mM MgCl_2 at pH 7.8 and 26 °C. **a**, **b**, Time courses of round 5 and 12 pools, respectively. Arrowheads mark 3' cleavage products; the smaller 5' fragments are not shown. **c**, Cleavage sites of round 5–16 RNAs. End-labelled primers complementary to the 5' constant region were used to reverse transcribe pool RNAs after self-cleavage. The lengths of the RT products were determined by comparison with markers from a sequencing reaction run in the same gel.

activity first clearly emerged, many unique sequences were present, consistent with the high diversity of low-activity ribozymes previously observed by Tang and Breaker⁷.

Clones that contained the following consensus sequence 5'-CTGANGA...GAAA-3', which is characteristic of the hammerhead ribozyme^{8,9}, were rare in round 5, but increased in frequency in the round 7 and 9 sequences, and completely dominated the round 11 and 12 sequences (Table 1). Remarkably, three of the round 11 clones (not shown) and one round 12 clone contained two complete hammerhead motifs. Further evidence that clones containing the hammerhead consensus sequences do encode hammerhead ribozymes comes from examination of the aligned sequences (Fig. 3). The sequences 5' of the CTGANGA and 3' of the GAAA are complementary to the cleavage region of the pool, and if base paired would be consistent with the secondary structure of the hammerhead ribozyme (Fig. 3a) and the observed cleavage products (Fig. 2c). In most round 12 clones (Fig. 3b), the region between the CTGANGA and GAAA elements can form a stem-loop structure, which would correspond to stem-loop II of the hammerhead ribozyme. In some clones, this putative stem is capped by an extensive loop, whereas in others a compact stem-loop structure is present. We tested the activity of several clones from this round individually (Fig. 3); they ranged from 0.89 min⁻¹ (clone 6) to 0.083 min⁻¹ (clone 16), comparable

to the activities of natural hammerhead ribozymes.

On the basis of the sequence and structural features noted above, we consider the selected ribozymes to belong to the hammerhead class of self-cleaving ribozymes. However, a closer examination of the aligned sequences revealed an unexpected degree of structural variation in what are normally viewed as conserved elements of the hammerhead ribozyme. In addition, some of the other hammerhead ribozymes that we observed contain mismatched or surprisingly short base-paired stems. Our mutational analysis of these ribozymes strongly supports the idea that their core structure is that of a hammerhead ribozyme, and suggests that these deviations from the hammerhead consensus are compensated for by other regions of the selected sequence. The single non-hammerhead clone found in round 12 (clone 10, not shown) was highly active with a self-cleavage rate constant of 0.74 ± 0.014 min⁻¹. No obvious similarity could be seen between this ribozyme and any of the known natural self-cleaving RNAs.

Our experiments firmly suggest that the hammerhead ribozyme is the simplest RNA motif capable of catalysing RNA strand cleavage at rates between 0.1 and 1.0 min⁻¹, because this single motif dominated the selected population once those rates had been attained. Other RNA motifs that catalyse self-cleavage with similar rates must either be more complex (and thus less common in samples of random-sequence RNA), must amplify significantly less

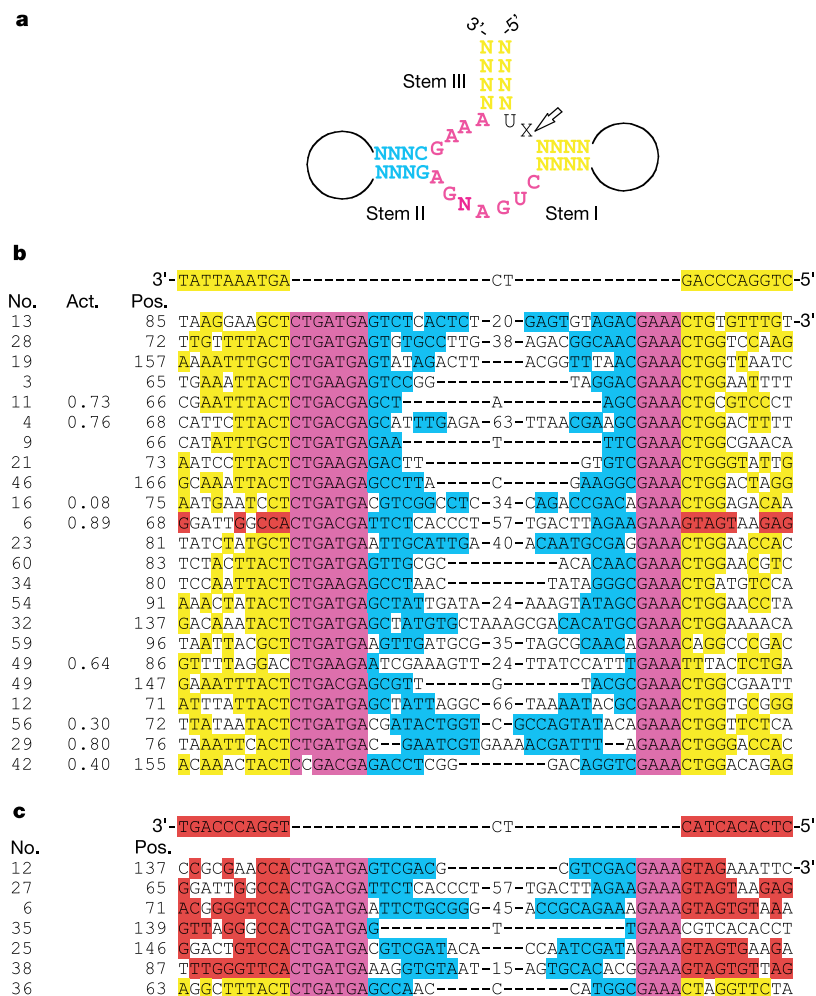


Figure 3 Alignments of unique hammerhead-type clones from rounds 12 and 16.

a, A standard hammerhead ribozyme¹⁶. X is A, U or C. **b**, **c**, Alignment of DNA sequences derived from the random region of the library, for clones from rounds 12 and 16, respectively, with substrate cleavage sites at the top. No., clone number; Act., self-

cleavage rates (min⁻¹); Pos., position of first base shown from 5' end of transcript. Colours represent the following: magenta, putative core hammerhead motifs; yellow, potential base pairing with the cleavage region shown at the top; red, potential base pairing with the alternative cleavage region; blue, potential stem II base pairing.

Table 1 Appearance of hammerhead clones during selection

Round	Frequency*	Activity† (min ⁻¹)	No.‡	Progenitors§
5	0.02	0.0026 ± 0.000057	40	1/39
7	0.17	0.026 ± 0.0021	29	20/5
9	0.67	0.086 ± 0.0022	33	23/8
11	1.00	0.166 ± 0.0053	30	17/0
12	0.98	0.322 ± 0.014	55	22/1
13	—	0.655 ± 0.022	—	—
		0.0381 ± 0.0031		
14	0.93	0.835 ± 0.054	30	19/1
		0.050 ± 0.00098		
16	0.77	0.289 ± 0.031	30	7/2
		0.0649 ± 0.0069		

* Observed frequency of hammerhead-containing clones.

† Estimated pool activity (mean ± s.d.).

‡ Number of sequenced clones.

§ Number of distinct, independent sequences with/without hammerhead motif.

|| Activity in 0.5 mM MgCl₂ at pH 7.2.

efficiently by the RT–PCR procedure we used, or must be restricted to specific substrate sequences not included in our selection. Our results are consistent with those of Tang and Breaker⁷, who observed many distinct ribozymes with activities lower than 0.1 min⁻¹. In contrast, another selection experiment¹⁰, in which the random-sequence region contained only 18 nucleotides, yielded only hammerhead ribozymes at both low and high activities. However, the stringent size constraint on the random-sequence region, and the constraints placed on cleavage site recognition, may have ruled out the appearance of other ribozymes. There are several natural classes of self-cleaving ribozymes other than the hammerhead, including the hairpin ribozyme¹¹, the hepatitis delta ribozyme (HDV)^{12,13}, and the *Neurospora* VS motif¹⁴. We wondered why none of these alternative highly active ribozymes was observed in our selection. These ribozymes are all significantly more complex than the hammerhead ribozyme, and are thus expected to occur much less frequently in selection experiments, in the absence of selective factors other than rate of cleavage. Thus, if selective pressure requires only that self-cleavage activity be in the range of 0.1–1.0 min⁻¹, the expected result, both *in vitro* and in nature, is the emergence of the hammerhead ribozyme as the most probable solution. Our results show that, despite the dominance of contingency (historical accident) in some recent discussions of evolutionary mechanisms¹⁵, purely chemical constraints (that is, the ability of only certain sequences to carry out particular functions) can lead to the repeated evolution of the same macromolecular structures. □

Methods

Selection procedure

Pool RNA dissolved in 10 mM Tris–HCl buffer at pH 8.0 and 1 mM EDTA (TE) was heated to 85 °C then cooled to ambient temperature. Cleavage reactions were started by adding pool RNA to reaction buffer such that the final composition was 1 μM RNA, 5 mM MgCl₂, 50 mM Tris–HCl at pH 7.8, 10 mM DTT and 1 U μl⁻¹ RNasin. Reaction times were 17–20 h at 26 °C in rounds 1–5 and in the subsequent rounds were adjusted to produce 1% cleavage. Reactions were stopped by addition of two volumes of 8 M urea, 2 mM Tris–HCl at pH 7.5 and 75 mM EDTA. Samples were resolved by electrophoresis in 8% denaturing polyacrylamide gels alongside a marker RNA to identify the position of the cleaved products; this region was excised and the RNA was isolated. RNAs were reverse transcribed using SuperScript II (Gibco–BRL), amplified by PCR and transcribed with T7 RNA polymerase (0.5 mM nucleoside triphosphates, 40 mM Tris–HCl at pH 7.9, 6 mM MgCl₂, 2 mM spermidine, 10 mM dithiothreitol and 1 U μl⁻¹ RNasin) in the presence of 50 μM of the blocking oligonucleotide (5′-TACTGACTGGGTCCAG-3′). After transcription, RNAs were purified by denaturing polyacrylamide gels, eluted, ethanol precipitated, dissolved in TE, and stored at –80 °C.

Cleavage site determination

RNAs were denatured with heat, cooled to ambient temperature, incubated in cleavage buffer, and then reverse transcribed with a 5′-³²P-labelled primer (5′-CGTTGTGGTCTATC-3′). Samples mixed with formamide stop buffer were analysed by electrophoresis in 12% denaturing polyacrylamide gels.

Received 25 May; accepted 7 August 2001.

- Pabon-Pena, L. M., Zhang, Y. & Epstein, L. M. Newt satellite 2 transcripts self-cleave by using an extended hammerhead structure. *Mol. Cell. Biol.* **11**, 6109–6115 (1991).
- Zhang, Y. & Epstein, L. M. Cloning and characterization of extended hammerheads from a diverse set of caudate amphibians. *Gene* **172**, 183–190 (1996).
- Ferbyre, G., Smith, J. M. & Cedergren, R. Schistosoma satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18**, 3880–3888 (1998).
- Rojas, A. A. *et al.* Hammerhead-mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets. *Nucleic Acids Res.* **28**, 4037–4043 (2000).
- Diener, T. O. Circular RNAs: relics of precellular evolution? *Proc. Natl Acad. Sci. USA* **86**, 9370–9374 (1989).
- Williams, K. P., Ciafre, S. & Tocchini-Valentini, G. P. Selection of novel Mg²⁺-dependent self-cleaving ribozymes. *EMBO J.* **14**, 4551–4557 (1995).
- Tang, J. & Breaker, R. R. Structural diversity of self-cleaving ribozymes. *Proc. Natl Acad. Sci. USA* **97**, 5784–5789 (2000).
- Thomson, J. B., Tuschl, T. & Eckstein, F. in *Catalytic RNA* (eds Eckstein, F. & Lilley, D. M. J.) 173–196 (Springer, Heidelberg, 1997).
- Ruffner, D. E., Stormo, G. D. & Uhlenbeck, O. C. Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry* **29**, 10695–10702 (1990).
- Conaty, J., Hendry, P. & Lockett, T. Selected classes of minimised hammerhead ribozyme have very high cleavage rates at low Mg²⁺ concentration. *Nucleic Acids Res.* **27**, 2400–2407 (1999).
- Hampel, N. in *Progress in Nucleic Acids Research and Molecular Biology* (ed. Moldave, K.) 1–38 (Academic, London, 1998).
- Been, M. D. *Cis-* and *trans-*acting ribozymes from a human pathogen, hepatitis delta virus. *Trends Biochem. Sci.* **19**, 251–256 (1994).
- Thill, G., Vasseur, M. & Tanner, N. K. Structural and sequence elements required for the self-cleavage activity of the hepatitis delta virus ribozyme. *Biochemistry* **32**, 4254–4262 (1993).
- Beattie, T. L., Olive, J. E. & Collins, R. A. A secondary-structure model for the self-cleaving region of *Neurospora* VS RNA. *Proc. Natl Acad. Sci. USA* **92**, 4686–4690 (1995).
- Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* 292–323 (Norton, New York, 1989).
- Long, D. M. & Uhlenbeck, O. C. Self-cleaving catalytic RNA. *FASEB J.* **7**, 25–30 (1993).

Acknowledgements

We thank P. Svec for assistance in cloning and sequencing, and members of the laboratory for their comments on the manuscript. This work was supported in part by a grant from the National Institutes of Health. J.W.S. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to J.W.S. (e-mail: szostak@molbio.mgh.harvard.edu).

corrections

Retrotransposition of a bacterial group II intron

Benoit Cousineau, Stacey Lawrence, Dorie Smith & Marlene Belfort

Nature **404**, 1018–1021 (2000).

In this Letter, we demonstrated active group II intron transposition to multiple ectopic sites and showed that movement to these new sites occurs via an RNA intermediate in a process termed retrotransposition. Of interest was whether retrotransposition occurs by reverse splicing of the intron RNA into RNA, single-stranded DNA and/or double-stranded DNA targets. Our results suggested that retrotransposition occurs predominantly by the intron reverse-splicing into an RNA target. However, subsequent experiments using a different intron donor have indicated that the selection method for retrotransposition events can strongly bias the interpretation of pathway (K. Ichyanagi, A. Beauregard and M.B., unpublished results). The new data do not point to RNA-targeting as the principal means for intron retrotransposition, and DNA-based pathways must be considered as playing an important role. □

Crystal structures of SarA, a pleiotropic regulator of virulence genes in *S. aureus*

Maria A. Schumacher, Barry K. Hurlburt & Richard G. Brennan

Nature **409**, 215–219 (2001).

In light of the recent X-ray structure determination of the SarR-maltose binding fusion protein by Zhang *et al.*¹, we re-examined our crystal structure determinations of the *Staphylococcus aureus* transcription regulator, SarA, which were reported in the 11 January 2001 issue of *Nature*. Although suggested homologues (27% identity), the SarA and SarR structures are significantly dissimilar. We now believe that our structures of the apo and DNA-bound forms of SarA may harbour some anomalies.

We are performing a comprehensive examination of the structure and function relationships of a global virulence gene regulator in *S. aureus*, SarA. The full-length, recombinant protein was expressed in *E. coli* and purified to apparent homogeneity. This protein, whether generated in the laboratory of B.K.H. or that of R.G.B., behaved as expected. It bound with high affinity to *cis* regulatory sequences upstream of virulence genes previously reported to be controlled by SarA. These data were reported in refs 2 and 3.

To investigate why the SarA and SarR structures are different, we first reassessed our structure-determination processes. The SIR and MAD (as well as averaged)-derived phases that were used to calculate our SarA–DNA complex structure resulted in electron density maps that showed consistent secondary structure features different from those of SarR. In addition, MAD data for one of the ‘apo’ SarA crystals revealed similar features to the DNA-bound form of SarA. These results suggested to us that we were working with a highly flexible molecule that could undergo remarkable structural changes. One point is the non-standard, although we feel correct, manner by which we calculated the value of R_{free} , that is, both structures were not refined with CNS and therefore we may have erred in the use of this quality-control standard. It remains a possibility that our intensity data were simply not of high enough quality to discern the structure of Zhang *et al.*¹. Although attempted multiple times, however, molecular replacement experiments to solve the apo SarA and the SarA–DNA protein structures, using our intensity data and SarR as the search model, have failed. Moreover, our heavy-atom sites are inconsistent with their SarR location. We are left with the possibility that in the absence of a carrier protein, such as MBP, or cognate DNA of sufficient length, we have crystallized a protein with some anomalously folded regions, which in our structures display a great deal of conformational flexibility. In support of this are the facts that there is a remarkable change in space group (from $P2_12_12_1$ to $P2_12_12$) and *c* cell edge (from 141 Å to 27 Å) upon freezing; all cryocooled crystals were non-isomorphous; our protein becomes inactive over time and degrades; our SarA–DNA complex reveals only nonspecific contacts; and there is an unprecedented change in protein conformation upon ligand binding.

To test the validity of our structures and to assess the importance of residues shown in the structures as important for function, we generated mutants of SarA that should result in aberrant activity. Using both an *in vivo* assay for virulence gene regulation and an *in vitro* DNA-binding assay for SarA function, most of the mutants have significantly altered activity (K. Sterba, M. S. Smeltzer and

B.K.H., unpublished results). At this point, we do not know the significance and origin of the differences between the reported structures of SarA and SarR. □

1. Liu, Y. *et al.* Crystal structure of the SarR protein from *Staphylococcus aureus*. *Proc. Natl Acad. Sci. USA* **98**, 6877–6882 (2001).
2. Rehtin, T. M. *et al.* Characterization of the SarA virulence gene regulator of *Staphylococcus aureus*. *Mol. Microbiol.* **33**, 307–316 (2001).
3. Blevins, J. S., Gillaspay, A. F., Rehtin, T. M., Hurlburt, B. K. & Smeltzer, M. S. The staphylococcal accessory regulator (sar) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (*cna*) in an *agr*-independent manner. *Mol. Microbiol.* **33**, 317–326 (2001).

Biodegradation of oil in uplifted basins prevented by deep-burial sterilization

A. Wilhelms, S. R. Larter, I. Head, P. Farrimond, R. di-Primio & C. Zwach

Nature **411**, 1034–1037 (2001).

Figure 2 inadvertently contained plotting errors for oils from the Tampen Spur and Viking Graben data sets and the plotting of pristane/*n*-heptadecane for the Barents Sea data set rather than pristane/*n*-heptadecane as described. The corrected figure is shown below with all the offshore data now correctly plotted as depth from sediment surface. The shallowest Barents Sea sample plots with pristane/*n*-heptadecane slightly greater than 1, but alkane, isotope and light hydrocarbon distributions suggest that this oil is not biodegraded. Reservoirs in the Barents Sea cooler than 80 °C at present are found at depths shallower than 2,000 m. Our conclusions remain unchanged. □

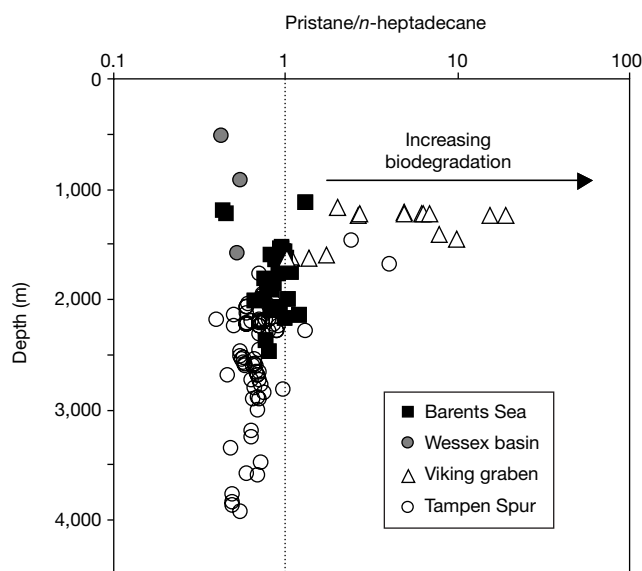


Figure 2 Degree of oil biodegradation, assessed by the pristane/*n*-heptadecane ratio, as a function of current reservoir depth below sediment surface for oils from four different settings.

nature insight

Stem cells



Cover illustration

Repeated images of coloured scanning electron micrograph of bone marrow stem cells. (Courtesy of Eye of Science/SPL.)

Stem cells are truly remarkable. They bridge the gulf between the fertilized egg that is our origin and the architecture that we become. They supply the cells that construct our adult bodies and, as we age, replenish worn out, damaged and diseased tissues. They renew themselves, resisting the powerful pull towards differentiation that overcomes more prosaic cells. And depending on the source, they have the potential to form one, many or all cell types of an organism.

Stem cell research has a history of more than 20 years, and has made some outstanding contributions to our understanding of haematopoiesis and mouse embryology. But the field has been transformed in the past few years by successes achieved in culturing human embryonic stem cells, the building blocks for every tissue we comprise, and in manipulating their differentiation *in vitro*. More recently, excitement has been fuelled by the controversial evidence that adult stem cells have a much higher degree of developmental potential than was previously imagined.

Scientists now face the formidable task of justifying all the attention by bringing stem cell therapies to the clinic. But this will demand a better understanding of stem cells at the molecular level and of how they behave in their biological context. The factors that maintain stem cells in a multipotent, proliferative state, or drive them to create differentiated daughter cells *in vitro* or *in vivo*, must be identified.

In commissioning this Insight, we have asked top researchers to cut through the hype and explain the fundamentals of stem cell biology as they see it today. We are indebted to all the contributors, and we are pleased to acknowledge the financial support of Osiris Therapeutics in producing this Insight. As always, *Nature* carries sole responsibility for the editorial content and peer review process. We hope that scientists, clinicians and general readers alike will find these articles both enlightening and thought provoking.

Natalie DeWitt Senior Editor

overview:

- 88 **The future for stem cell research**
R. Lovell-Badge

review articles:

- 92 **The end of the beginning for pluripotent stem cells**
P. J. Donovan
& J. Gearhart

- 98 **Stem cells find their niche**
A. Spradling,
D. Drummond-Barbosa
& T. Kai

- 105 **Stem cells, cancer, and cancer stem cells**
T. Reya, S. J. Morrison,
M. F. Clarke
& I. L. Weissman

- 112 **The development of neural stem cells**
S. Temple

progress:

- 118 **Stem cells in tissue engineering**
P. Bianco & P. G. Robey

review articles:

- 122 **Reprogramming of genome function through epigenetic inheritance**
M. A. Surani

commentary:

- 129 **Ethical and social considerations of stem cell research**
A. McLaren

The future for stem cell research

Robin Lovell-Badge

*Division of Developmental Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
(e-mail: rlovell@nimr.mrc.ac.uk)*

Stem cells have offered much hope by promising to greatly extend the numbers and range of patients who could benefit from transplants, and to provide cell replacement therapy to treat debilitating diseases such as diabetes, Parkinson's and Huntington's disease. The issue of stem cell research is politically charged, prompting biologists to begin engaging in ethical debates, and generating in the general public an unusually high level of interest in this aspect of biology. But excitement notwithstanding, there is a long way to go in basic research before new therapies will be established, and now the pressure is on for scientists and clinicians to deliver.

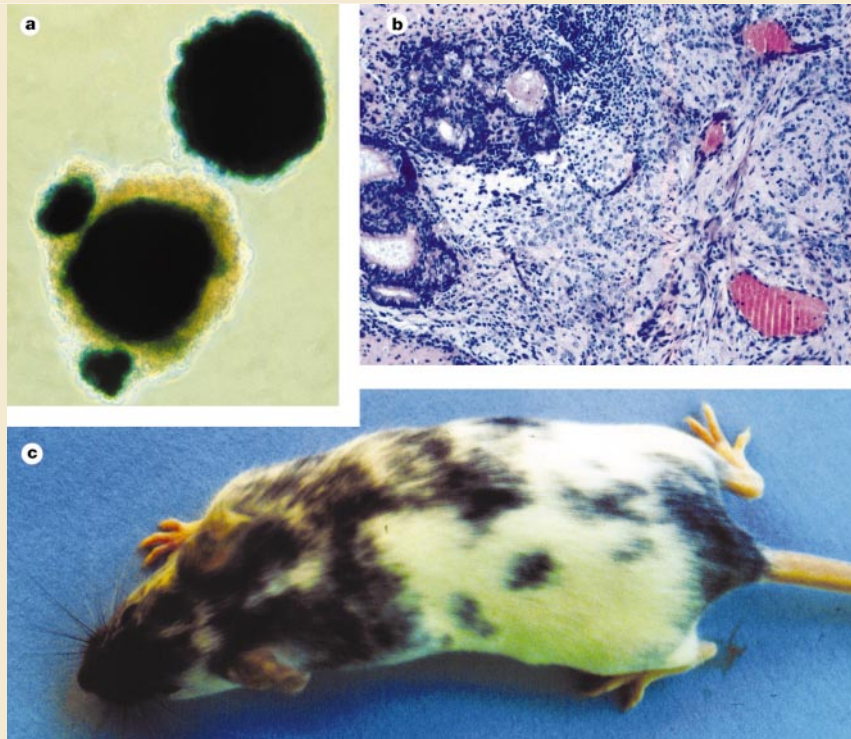
Scientists and clinicians have been beavering away in the background on stem cell research for years, but it took two breakthroughs — animal cloning by Ian Wilmut, Keith Campbell and colleagues¹ and the derivation of pluripotent human embryonic stem (ES) cells by Jamie Thomson and co-workers² — to really shake things up. The first showed that an adult cell nucleus can be reprogrammed to produce an entire animal, a dramatic demonstration of hidden potential, and the second provided a possible source of cells for cell-based therapies for many human diseases. Even more intriguing was the notion of combining the two in the process of 'therapeutic cloning', whereby ES cells could be (in theory anyway) derived from a patient's cell, thus avoiding problems of immune rejection, and then coaxed into forming large quantities of the cell types needed for a cure. The derivation of human embryonic germ (EG) cell lines from aborted fetuses, with a similar potential to ES cells, by

John Gearhart's group³, provided an alternative route to obtain cells for treating patients, avoiding the need to use early embryos. But then, more recently, adult stem cells have leapt into prominence with many claims that they are more plastic than previously thought, and can in some circumstances contribute to cell types very different from those in their tissue of origin. This has suggested that patient-specific stem cell therapy can be carried out without cloning.

As is evident from reading the articles comprising this Insight, stem cell research has opened for exploration a vast new terrain of basic biology. Fundamental questions remain, such as what defines a stem cell in molecular terms, what signalling events control stem cell differentiation, and what does it mean to reprogramme a cell? Scientists are now reconsidering the very definition of a stem cell, with the attendant notions of lineage restriction and the permanency of the differentiated cell state. Although this uncertainty is unnerving, in the end we will profit from a better

Figure 1 Pluripotency of mouse embryonic stem (ES) cells.

a, Aggregates of mouse ES cells forming embryoid bodies. The dark staining shows expression of Sox2 in the less differentiated cells, whereas the ring of differentiated endoderm is unstained. (Image courtesy of A. Avilion.) **b**, Histological section of a teratocarcinoma derived from mouse ES cells. Many different cell types are found, all formed from the ES cells, including representatives of all three germ layers. (Image courtesy of M. Parsons.) **c**, The ultimate test of pluripotentiality: a chimaera made by injecting ES cells into a blastocyst. The pigmented areas reveal the contribution of ES cell derivatives to the skin, but all tissues are composed of a mixture of ES and host embryo derivatives. Even a single ES cell can give a chimaera like this.



understanding of how stem cells work as well as how to identify, isolate and maintain them, and trigger their differentiation along useful directions.

We would love there to be common rules, but given the multitude of stem cell types, this may be asking too much. Some types of stem cells, such as bone marrow and skin, have already been used in therapies, and others are effectively being used in trials, including fetal midbrain cells for Parkinson's disease, and pancreatic duct cells for diabetes. But all have different properties reflecting the normal requirement for the tissue to which they belong. Some are easy to identify and to grow; others remain a mystery. Paradoxically, one of the easiest types to grow and expand in culture are ES cells, which are those with the greatest potential. This is one reason why it is so important to learn how to control their differentiation and to be able to select specific cell types to be used for therapies⁴. It is also crucial to understand the nature of pluripotentiality, one of the defining properties of ES cells (Fig. 1), discussed at length in the articles by Azim Surani (pages 122–128) and by Peter Donovan and John Gearhart (pages 92–97).

In political terms, the use of early embryos to derive ES cells has stimulated intense debate among professional ethicists, moral philosophers, religious leaders and scientists. Most ethicists believe that human embryo research raises no new ethical problems, at least none that were not already considered with respect to methods of assisted conception, such as *in vitro* fertilization. If ES cells turn out to be the best route to cure a particular disease, then many would argue that it would be morally wrong not to use embryos that would otherwise be discarded. Anne McLaren discusses the ethical and social considerations of stem cell research on pages 129–131.

Realizing potential

The story of ES cells starts from the study of teratocarcinomas, first worked on in the mouse by Leroy Stevens, and discussed in this issue by Donovan and Gearhart. These are complex tumours containing a mix of differentiated cell types and a population of undifferentiated cells, termed embryonal carcinoma (EC) cells. The latter were shown to give rise to cell types from all three classical embryonic germ layers: ectoderm, mesoderm and endoderm⁵ (Fig. 1b). The idea that such pluripotent EC cells might provide a source of cells for therapy was therefore around since at least the mid-1970s and encouraged much work on human EC cells. But they never seemed ideal because they were aneuploid and had come from tumours. It was known that teratocarcinomas could be made experimentally by transplanting early postimplantation mouse embryos to ectopic sites, so it seemed natural to try to avoid the tumour step by putting early embryos directly in culture. An understanding of the biology of EC cells and early embryos led Martin Evans and Matt Kaufman to put blastocysts that had been kept in diapause (or delayed implantation) into culture conditions that had been optimized for the best available EC cell lines. The successful derivation of ES cells that followed⁶, and the demonstration that they had a normal karyotype and were truly pluripotent⁷, has of course had ramifications far outside even the important area of stem cell biology.

ES cells are known to correspond roughly to cells of the early epiblast, but they are not identical to them. Mouse ES cells depend on LIF (leukaemia inhibitory factor) to remain undifferentiated, whereas embryos mutant in LIF or its receptors show normal development. Nevertheless, there is a requirement for LIF signalling by the inner-cell-mass cells of blastocysts in diapause⁸. Perhaps these cells are closer to ES cells, even though they are essentially non-dividing cells. On the other hand, it is known from many systems that stem cells are often quiescent. Perhaps this will also turn out to be a theme — stem cells maintained *in vitro* may rarely correspond to their counterparts *in vivo*.

Cells fitting the classic definition of stem cells also exist in the adult, where they function to replace cells lost from normal turnover or damage. Through asymmetric divisions, adult stem cells produce

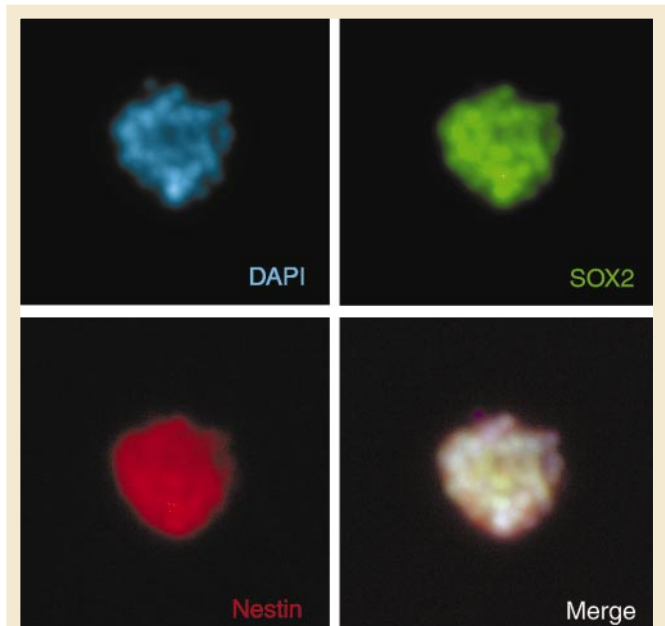


Figure 2 A small primary neurosphere, obtained from one or a few cells from the dorsal telencephalon of a 14.5-d.p.c. mouse embryo, which has been grown in culture for 21 days. This is stained with DAPI (blue) to reveal nuclei and with both SOX2 (green) and Nestin (red) to reveal neural progenitor cells. Many, if not all, of these cells have properties of neural stem cells. (Figure courtesy of E. Remboutsika.)

just the right number of differentiated cells, as well as daughter stem cells with properties identical to those from which they arose. This is true self-renewal. In contrast, in the embryo, although many cells undergo a similar asymmetric division with respect to daughter cell fate, this does not necessarily involve exact self-renewal. The daughter stem cells progressively become restricted in the variety of differentiated progeny they can produce, thereby losing developmental potential. This can occur rapidly, as in the transition between the inner cell mass and epiblast, or gradually, as in the developing central nervous system (CNS), as discussed by Sally Temple on pages 112–117. In considering the range of cell types to which a stem cell can give rise, Temple highlights the importance of the cell's environment and the increasing complexity of the embryo as it develops, as well as its history — cells isolated from different parts of the developing CNS remember from where they have come, even after many cell divisions *in vitro* (Fig. 2).

This notion of history versus environment has been discussed eloquently by David Anderson elsewhere⁹. As he points out, without history (or lineage) providing some hard-wiring, in combination with an organized body plan to put signals in the right place, it would be difficult to build an embryo, whereas environment alone must dictate what a stem cell does in the adult. But, seemingly to the surprise of many an embryologist, sometimes environment is able to overcome any amount of careful upbringing and historical propaganda. This is evident in recent data suggesting that adult stem cells isolated from one tissue can give rise to cell types characteristic of another. They can even cross over from one classic germ layer to another¹⁰. There are many dramatic examples of this, some mentioned in the following articles. But perhaps the most extreme example is that from Frisen and colleagues showing that adult CNS stem cells can contribute to cell types of all three primary germ layers after being introduced into blastocysts, where they behave something like ES cells¹¹.

Home alone

However, one important caveat with many of the recent experiments looking at the potential of stem cells is that they are rarely performed with single cells. Picking single EC or ES cells and showing that they

can give rise to a wide range of cell types *in vitro* or *in vivo* was the only real proof that they were truly pluripotent stem cells^{5,7}. If, on the other hand, one begins with a population of cells, then that is all one is testing — many different cells, perhaps each with a unique potential. The possibility of contamination of one stem cell type with another has been stressed by Tannishtha Reya, Irving Weissman and colleagues (pages 105–111), who caution that haematopoietic stem cells (HSCs) could well be distributed throughout the body. The presence of differentiated cells among the stem cells might also affect results, but perhaps in the other direction. In Frisen's experiments, neurospheres were able to contribute to normal development to make chimaeric embryos in only 1% of attempts¹¹. This could mean that the potential to do so was found only in rare neural stem cells. Alternatively, if differentiated cells tend to be present in the neurospheres, these could exert an inhibitory influence on the stem cells, in most cases preventing them from fulfilling their true potential. These are all problems that could impinge on the usefulness of stem cells in the clinic.

It is therefore of no surprise that the importance of testing the fate of single cells is stressed by several authors, notably by Alan Spradling and colleagues (pages 98–104) and Sally Temple (pages 112–117). Even if we set aside problems of contamination, it would be useful to know if some unusual potential, such as neural stem cells contributing to muscle, is a property of specific rare cells in the population or whether each cell has a low probability of differentiating in this direction. And even if the population of cells is clonally derived, there could still be variation, which could be explained if stochastic epigenetic changes affected the activity of certain critical genes. In fact, this could well be the explanation for the variable patterns of imprinted gene expression and phenotypes seen in cloned mice made using donor nuclei from ES cells¹². But it is not always easy to test single cells. It has been attempted for bone-marrow stem cells, although a rather complex procedure had to be adopted to overcome the problem in identifying them in the first place¹³. We clearly need better markers for stem cells and efficient ways to test their potential both *in vitro* and *in vivo*.

Home truths

Spradling and colleagues highlight the importance of the stem cell niche — essentially the environment in which stem cells exist and which contributes to their self-renewal and differentiation. Cell–cell or cell–basement membrane contacts could provide a simple means by which asymmetric cell division can lead to these alternate fates. However, more diffuse signalling, such as by growth factors, could also be used to regulate the proportion of stem versus differentiated cells and rates of proliferation. This leads to the notion of two types of niche, lineage based and population based, but it is stressed that in many cases both may operate together.

They also suggest a rigorous way in which a stem cell niche can be identified, by showing that a single stem cell can repopulate an empty niche. But there is a caveat to this. The proper functioning of niches is likely to depend on the continued presence of stem cells and empty niches are likely to lose their abilities and degenerate. Testes devoid of germ cells eventually lose Sertoli cells (P. Burgoyne, personal communication). It is therefore not surprising if they also lose the ability to support newly transplanted spermatogonia (germ stem cells)¹⁴. This may be a problem that needs to be faced with stem cell-based therapies, that is, it may be necessary to rejuvenate or replace the niche as well as the stem cells. This area is covered by Paolo Bianco and Pamela Robey on pages 118–121, in an article which places stem cell biology in the context of tissue engineering, and highlights some of the many practical problems to be encountered in achieving cell-based therapies.

Plus ça change

Other types of cells, notably HSCs, seem to differentiate in the absence of a niche or the differentiating cell population, changing from long-term, self-renewing true stem cells to transient amplifying cells in a process that is apparently resistant to environmental influences, as discussed by Irving Weissman and colleagues. They also discuss the

Box 1

A few definitions

Blastocyst. A mammalian embryo before implantation. Blastocysts consist of outer trophoblast cells, which allow the embryo to implant, surrounding the inner cell mass, within which are found the pluripotent epiblast cells.

Cell fate. What a cell can do in either its natural location in the embryo or in an ectopic site. This is usually determined by marking the cell in as neutral a way as possible.

Committed. Used to describe cells whose fate is already determined along a particular path of differentiation — at least within the bounds of the experimental assay.

Epigenetic. Heritable but reversible changes in gene function without changes in DNA sequence. Usually involves chemical modifications of DNA or the binding of specific proteins to DNA sequences. Epigenesis is a related term used to describe formation of new structures during embryonic development.

Epigenetic asymmetry. Two genomes with identical DNA sequences, but with different forms of reversible but heritable modifications that regulate gene expression and repression.

Imprinted gene. A gene that is marked in the germ line; this denotes its maternal or paternal origin and influences its expression in the developing embryo.

Lineage. The natural progression from an immature cell type to one or more differentiated cell types.

Lineage restriction. The inability of one lineage to give cell types of another, that is, to cross lineage boundaries.

Molecular memory. The inheritance of specific pattern of gene expression, which persists in daughter cells from generation to generation.

Multipotent. Able to give rise to more than one differentiated cell type.

Plasticity. The ability to cross lineage boundaries.

Pluripotent. Able to give rise to all cell types found in the embryo and adult animal.

Progenitor cells. These can include both stem cells and transient amplifying cells, or even cells that are well on the way to becoming differentiated.

Stem cell. Generally used to describe a cell that is capable of both self-renewal and differentiation. However, the term has been used in many other contexts. (See ref. 27, pages 1–16 for more details, and the entire volume for many current ideas on stem cell biology.)

Therapeutic cloning. Reprogramming the nucleus of an adult cell through transfer into the cytoplasm of an enucleated oocyte. This is sometimes referred to as 'cell-nuclear replacement' or 'somatic cell nuclear transfer'.

Totipotent. Able to give rise to all cell types. In mammals, only the fertilized egg and early cleavage stage blastomeres are truly totipotent. Cells of the inner cell mass and ES cells are unable to differentiate into cells of the trophoblast lineage.

Transient amplifying cells. Progeny of stem cells that undergo replication, but are not able to self-renew and eventually give only one or more differentiated cell types.

relationship between stem cells and cancer cells, and the evidence that mutations 'hijack' the self-renewing population of stem cells. They note that tumours are often a mix of cell types, with teratocarcinomas and Wilms' tumours being the obvious cases. Even when cell morphologies are similar, often only a small proportion of the cells are tumorigenic. But ES cells seem to break all the rules, as every undifferentiated cell is tumorigenic, without any mutations¹⁵. We clearly need to understand more about the relationship between cancer cells and stem cells, especially if we are to use the latter for therapies.

What are the molecular correlates of changes in stem cell potential, whether due to temporal, positional or lineage restrictions?

Some of these issues are raised by Azim Surani on pages 122–128. Presumably there are changes in the transcriptional activity of important genes within the stem cells. But are there any rules we can understand that might prove useful? Should we be looking for changes in the activation or repression of specific genes, or for changes in chromatin through the action of, for example, members of the polycomb and trithorax groups of factors, or for changes in histone modifications (acetylation and methylation) or methylation of DNA? No doubt it will be a combination of several of these processes acting on particularly critical genes. These genes may be of two types: those that confer stem cell properties, such as allowing self-renewal and maintaining potential, and those that give them a certain 'flavour', in other words, determining the range of differentiated progeny they can give rise to. If we can find the most critical of these genes then perhaps we could achieve the ambitious aim of reversing the process, turning differentiated cells directly into stem cells, which could then be used to cure disease or trauma. Both Spradling and colleagues and Temple mention the importance of identifying stem cell genes, but so far we know of very few: perhaps *piwi* as mentioned by the former, and *Sox2* and *Oct4* in early mouse development^{4,16}.

As mentioned above, it was a surprise to many that embryological rules seem to be broken by recent experiments on stem cells. One such rule is that of lineage restriction: cells derived from one of the three primary germ layers that arise at gastrulation — ectoderm, mesoderm and endoderm — should not be able to form cell types characteristic of another. The breaking of this rule has now been reported many times in experiments on stem cells^{11,13,17}. But, in fact we have known natural examples of this for a long time. Neural crest is ectodermal in origin, but it gives rise to tissues, such as cartilage and muscle during normal development of the head, that are indistinguishable from those that have arisen from mesoderm formed via gastrulation¹⁸. There are also several examples of transdifferentiation, such as within the eye; these are situations where cellular reprogramming occurs naturally, and examples have been studied for many years¹⁹.

Another rule of the embryo is the progression from an undifferentiated to more differentiated state. This rule can also be violated, as demonstrated recently by Kondo and Raff, who showed that O2A oligodendrocyte precursors can be induced by specific culture conditions to revert to neural stem cell-like cells²⁰. But this rule was also broken before, at least 9 years ago. The articles by Surani and by Donovan and Gearhart discuss EG cells: pluripotent cells similar to ES cells that are derived from primordial germ cells (PGCs) within (or migrating towards) the genital ridges. Germ cells are very specialized cells, and become more so as they differentiate into sperm and oocytes. They usually have to go through meiosis to realize their totipotency, and sperm achieve this only after fertilization and reprogramming of their DNA by egg cytoplasm. Thus, it is thought that the culture conditions used to establish EG cells effectively regress the PGCs to a more primitive state. Germ cells undergo a complex history during their formation, but by changing their environment in culture they can be reverted all the way back to cells with properties similar to those of the epiblast from which they arose^{21,22}.

Back to the future

It may also be useful to take an evolutionary approach in studying stem cells. There are known to be species differences in the way embryos are made, perhaps even for the CNS²³, suggesting that if it is possible to get to the same cell types by different routes, then we should not be so surprised if certain stem cells are able to do things that contravene established thinking. It is also interesting to ask why regeneration occurs readily in some vertebrates, but not others. This

includes the coordinated renewal of many cell types, such as with limb regeneration in newts²⁴, or cases where only a single cell type is involved. For example, cochlear hair cells of birds are renewed from adjacent cells with stem cell-like properties, whereas those of mammals are never replaced once lost^{25,26}. Perhaps this is because song is so important to bird society, although it is little consolation for the pheasant being shot at, that the human wielding the gun will probably go deaf.

The next few years will no doubt bring great advances in understanding stem cells at the molecular level. New techniques will aid in identifying critical genes involved in controlling their self-renewal and differentiation. Perhaps these will allow us to manipulate stem cells *in vivo* in a useful way. Similarly, genes involved in reprogramming will be found. So far the only even remotely reliable way of reprogramming an adult cell type into another is by transferring its nucleus into the cytoplasm of an oocyte. This presumably reflects the normal ability of the egg to reprogramme the incoming sperm DNA to behave like its own. But is this due to one or many cytoplasmic factors? Identifying these and understanding how can they restore totipotentiality will be a substantial but very worthwhile challenge. □

1. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
2. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
3. Shambloot, M. J. *et al.* Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl Acad. Sci. USA* **95**, 13726–13731 (1998).
4. Li, M., Pevny, L., Lovell-Badge, R. & Smith, A. Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr. Biol.* **8**, 971–974 (1998).
5. Martin, G. R. & Evans, M. J. The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**, 163–172 (1974).
6. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotent cells from mouse embryos. *Nature* **292**, 154–156 (1981).
7. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255–256 (1984).
8. Nichols, J., Chambers, I., Taga, T. & Smith, A. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**, 2333–2339 (2001).
9. Anderson, D. J. Stem cells and pattern formation in the nervous system: the possible versus the actual. *Neuron* **30**, 19–35 (2001).
10. Anderson, D. J., Gage, F. H. & Weissman, I. L. Can stem cells cross lineage boundaries? *Nature Med.* **7**, 393–395 (2001).
11. Clarke, D. L. *et al.* Generalized potential of adult neural stem cells. *Science* **288**, 1660–1663 (2000).
12. Humphreys, D. *et al.* Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95–97 (2001).
13. Krause, D. S. *et al.* Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. *Cell* **105**, 369–377 (2001).
14. Ohta, H., Yomogida, K., Dohmae, K. & Nishimune, Y. Regulation of proliferation and differentiation in spermatogonial stem cells: the role of c-kit and its ligand SCF. *Development* **127**, 2125–2131 (2000).
15. Langa, F. *et al.* Teratocarcinomas induced by embryonic stem (ES) cells lacking vimentin: an approach to study the role of vimentin in tumorigenesis. *J. Cell Sci.* **113**, 3463–3472 (2000).
16. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
17. Lumelsky, N. *et al.* Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science* **292**, 1389–1394 (2001).
18. Le Lievre, C. S. & Le Douarin, N. M. Mesenchymal derivatives of the neural crest: analysis of chimaeric quail and chick embryos. *J. Embryol. Exp. Morphol.* **34**, 125–154 (1975).
19. Eguchi, G. & Kodama, R. Transdifferentiation. *Curr. Opin. Cell Biol.* **5**, 1023–1028 (1993).
20. Kondo, T. & Raff, M. Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. *Science* **289**, 1754–1757 (2000).
21. Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotent embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847 (1992).
22. Resnick, J. L., Bixler, L. S., Cheng, L. & Donovan, P. J. Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550–551 (1992).
23. Streit, A. & Stern, C. D. Neural induction. A bird's eye view. *Trends Genet.* **15**, 20–24 (1999).
24. Brockes, J. P. Amphibian limb regeneration: rebuilding a complex structure. *Science* **276**, 81–87 (1997).
25. Roberson, D. F., Weisleder, P., Bohrer, P. S. & Rubel, E. W. Ongoing production of sensory cells in the vestibular epithelium of the chick. *Hear. Res.* **57**, 166–174 (1992).
26. Rivolta, M. N. *et al.* Auditory hair cell precursors immortalized from the mammalian inner ear. *Proc. R. Soc. Lond. B* **265**, 1595–1603 (1998).
27. Marshak, D. R., Gardner, R. L. & Gottleib, D. (eds) *Stem Cell Biology* (Cold Spring Harbor Laboratory Press, New York, 2001).

Acknowledgements

I thank M. A. Surani for contributions to the glossary in Box 1.

The end of the beginning for pluripotent stem cells

Peter J. Donovan* & John Gearhart†

*Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA (e-mail: pdonovan@lac.jci.tju.edu)

†The Institute of Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA (e-mail: gearhart@jhmi.edu)

Pluripotent stem cells can be expanded seemingly indefinitely in culture, maintain a normal karyotype and have the potential to generate any cell type in the body. As such they represent an incredible resource for the repair of diseased or damaged tissues in our bodies. These cells also promise to open a new window into the embryonic development of our species.

The new millennium promised to usher in the era of the human genome. So far, a different area of biology — stem cell biology — has captured both the scientific and international news headlines. Stem cells are unique cells that have the capacity for self-renewal and are capable of forming at least one, and sometimes many, specialized cell types. Such stem cells are present in many tissues of adult animals and are important in tissue repair and homeostasis. For example, spermatogonial stem cells in the testis are unipotent and produce only one type of differentiated cell, a spermatozoon; whereas haematopoietic stem cells are multipotent and produce erythrocytes and all the types of white blood cells. Pluripotent stem cells can give rise theoretically to every cell type in the animal body and are derived not from adult but rather from embryonic tissues. Three types of mammalian pluripotent stem cell lines have been isolated — embryonal carcinoma (EC) cells, the stem cells of testicular tumours; embryonic stem (ES) cells derived from pre-implantation embryos; and embryonic germ (EG) cells derived from primordial germ cells (PGCs) of the post-implantation embryo (Fig. 1).

If pluripotent stem cells derived from human embryos behave like their counterparts from mice, they could be used to treat a wide variety of human diseases, particularly those in which specific cell types (such as cardiomyocytes, dopaminergic neurons and β -islet cells) have been lost or disabled. But what is the reality? What are the important properties of pluripotent stem cells and how do they differ from adult stem cells? Do the results of studies in animal models suggest stem cells can be used to correct disease phenotypes? How close are we to taking stem cell-based treatments into the clinic? What problems must be surmounted? Recent advances in understanding the basic biology of pluripotent stem cells suggest that they may be as useful as predicted, but major hurdles remain to be overcome. Furthermore, the contribution that studies of these cells could have on understanding the developmental biology of our own species has been overshadowed by the hype surrounding the potential for stem cell-based therapies. For the first time we can begin to understand how cells of a human embryo grow and develop to form a new individual and how that process can sometimes go wrong. That opportunity comes with great responsibility, but also inspires great awe.

The science of pluripotency

Defining pluripotent stem cell lines

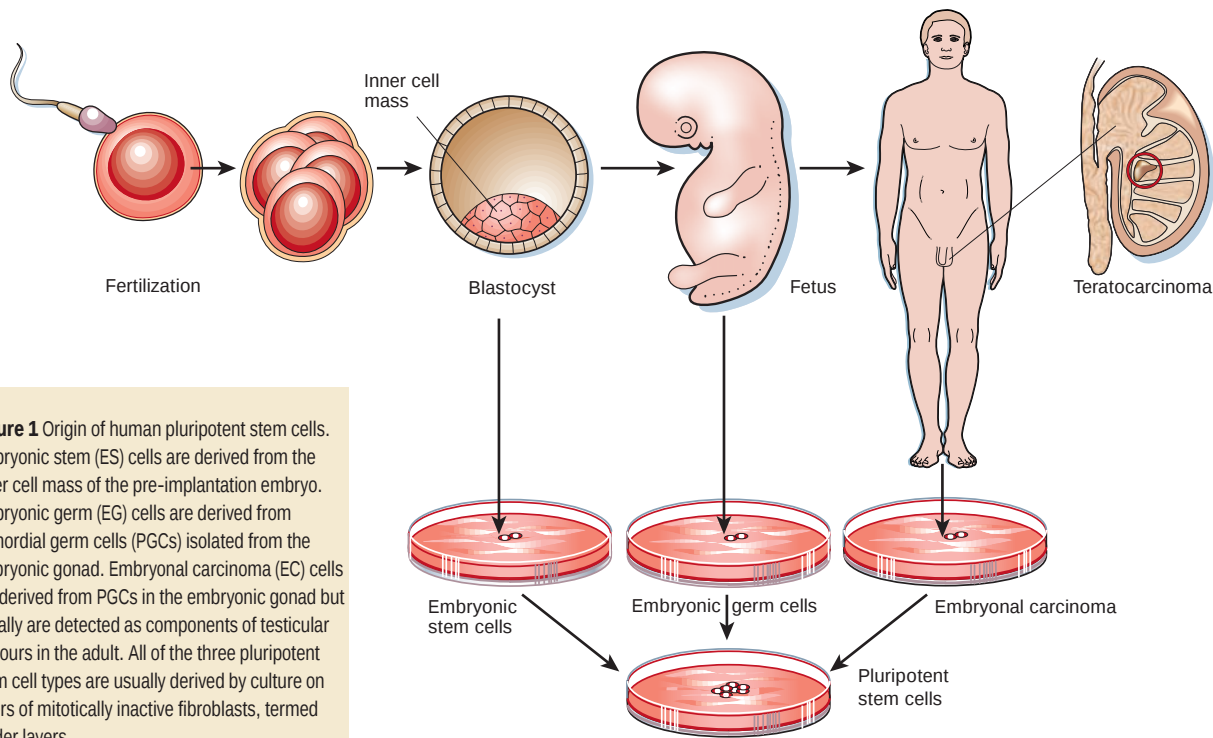
As a distinct cell type, the pluripotent stem cell was first recognized in teratocarcinomas. These are bizarre gonadal

tumours containing a wide array of tissues derived from the three primary germ layers that make up an embryo: the endoderm, mesoderm and ectoderm (refs 1,2, and see the overview in this issue by Lovell-Badge, pages 88–91). These tumours contain a large assortment of tissue types including cartilage, squamous epithelia, primitive neuroectoderm, ganglionic structures, muscle, bone and glandular epithelia. The differentiated cells of the tumour are formed from pluripotent EC cells present in the tumour, which themselves are derived from PGCs, the embryonic precursors of the gametes^{3,4}. EC cells are also one of the main components of human testicular germ cell tumours and, as in the mouse, evidence suggests that such tumours arise from PGCs⁵, although this has not been proven formally⁶. Cultured EC cell lines were derived by isolating EC cells from tumours and growing them in medium containing serum either in the presence or absence of a mitotically inactivated layer of fibroblasts, termed a feeder layer^{7–9}.

In contrast, ES cells are derived from the pluripotent inner cell mass (ICM) cells of the pre-implantation, blastocyst-stage embryo^{10,11}. Outgrowth cultures of blastocysts give rise to different types of colonies of cells, some of which have an undifferentiated phenotype. If these undifferentiated cells are sub-cultured onto feeder layers they can be expanded to form established ES cell lines that seem immortal.

And finally, EG cells are derived from cultured PGCs, the same cells from which EC cells are derived^{12,13}. PGCs isolated directly from the embryonic gonad onto feeder layers will, in the presence of serum and certain growth factors, form colonies of cells that seem morphologically indistinguishable from EC cells or ES cells grown on feeder layers (Fig. 1). Although EC, ES and EG cell lines have been isolated from mice and humans^{14–16}, only ES cells have been isolated from non-human primates^{17,18}.

The pluripotent stem cell lines have many attributes in common, with some exceptions of uncertain significance (Table 1). Some of the classical markers of these cells include an isozyme of alkaline phosphatase, the POU-domain transcription factor Oct4, high telomerase activity and a variety of cell-surface markers recognized by monoclonal antibodies to stage-specific embryonic antigens or to tumour-recognition antigens¹⁹. Although some of these markers are not unique to stem cells, they can nevertheless serve as reagents with which to physically separate pluripotent stem cells from their differentiated derivatives. The physiological significance of most of the markers is unclear, with the exception of Oct4. Compelling studies carried out in mouse EC, ES and EG cells, as well as in mouse embryos,



point to a critical role for Oct4 in the establishment and/or maintenance of pluripotent cells in a pluripotent state²⁰. Differentiation of pluripotent cells is associated with downregulation of Oct4 levels, and downregulation of the *Oct4* gene in ES cells or in mice results in the differentiation and loss of pluripotent cells^{21,22}.

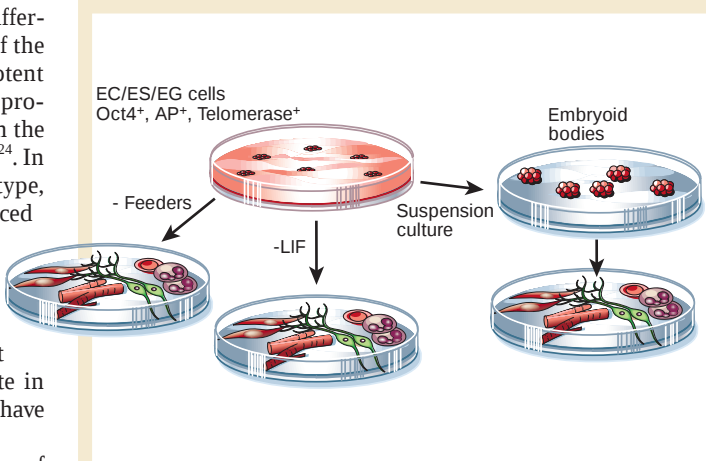
Developmental potential

The developmental potency of mouse pluripotent stem cells has been tested in three independent assays: *in vitro* differentiation in a Petri dish; differentiation into teratomas or teratocarcinomas when placed in adult histocompatible or immunosuppressed mice; and *in vivo* differentiation when introduced into the blastocoel cavity of a pre-implantation embryo. All of the pluripotent stem cells can differentiate *in vitro* into a wide variety of cell types representative of the three primary germ layers in the embryo (Fig. 2). When pluripotent stem cells are introduced into histocompatible or immunocompromised mice, they form tumours that are indistinguishable from the gonadal tumours from which EC cells were originally derived^{23,24}. In chimaeras, mouse ES and EG cells contribute to every cell type, including the germline^{25–28}. In contrast, murine EC cells introduced into embryos colonize most embryonic lineages, but generally do not colonize the germline, with one experimental exception^{29–32}. The inability of EC cells to form functional gametes most likely reflects their abnormal karyotype^{23,24}. Because of ethical concerns, non-human primate and human pluripotent stem cells have not been tested for their ability to participate in embryonic development *in vivo*, but in the other assays they behave identically to their murine counterparts^{14–16}.

The ability of pluripotent stem cells to give rise to a wide array of differentiated derivatives is, of course, the reason why they may be so useful for purposes of cell-based therapy. In the absence of factors that inhibit their differentiation (see Box 1), pluripotent stem cell differentiation has typically been directed by manipulating their environment by trial and error. This can be achieved by growing stem cells at high density, by growing them on different types of feeder cells, by addition of growth factors, or by growth on crude or defined extracellular matrix substrates (Fig. 2). In these conditions the types

Table 1 Comparisons among EC, ES and EG cells

	EC	ES	EG
Derivation	PGC	ICM	PGC
Karyotype	Heteroploid	Euploid	Euploid
Chimaera formation	Soma	Soma & Germ	Soma & Germ
Alkaline phosphatase	+	+	+
Telomerase	+	+	+
Oct4	+	+	+
<i>In vitro</i> differentiation	+	+	+



Box 1

Culture of pluripotent stem cells

An important property of pluripotent stem cells is their ability to divide symmetrically in culture and give rise to two daughter cells that are exact copies of the stem cell from which they were derived. This property allows pluripotent stem cells to be expanded in culture before induction of differentiation. Few of the factors that regulate self-renewal of pluripotent stem cells are known. Typically, pluripotent stem cell lines are isolated and maintained on mitotically inactive feeder layers of fibroblasts. EC cells maintained on feeder layers seem to retain developmental potency more readily than cells grown without feeders⁶⁶. Similar feeder-dependent culture conditions were used for the isolation of mouse and human ES and EG cells and such feeder layers proved critical to maintaining them in an undifferentiated state^{10–16}.

The requirement for feeder cells suggests that they provide a factor that suppresses the differentiation or promotes the self-renewal of pluripotent stem cells. An activity with these properties was originally termed differentiation-inhibiting activity (DIA)⁶⁷. DIA was found to be the same as leukaemia inhibitory factor (LIF), a member of the family of cytokines related to interleukin-6 (ref. 68).

For murine ES cells, LIF can replace the requirement for feeder cells⁶⁴. Importantly, activation of the signalling component of the LIF receptor, glycoprotein 130 (gp130), is both necessary and sufficient for inhibiting murine ES cell differentiation⁶⁹. A crucial downstream effector of gp130 is the signal transducer and activation of

transcription-3 (STAT3)^{70,71}. Other signalling molecules acting downstream of gp130, such as the mitogen-activated protein kinase, seem to actually inhibit ES cell self-renewal⁷².

Isolation of human ES cells requires feeder cells (and 20% fetal calf serum), but does not seem to require LIF^{14,16}. However, human ES cells grow on feeder cells that presumably produce multiple growth factors. Therefore, it is possible that some of the same downstream signalling molecules required for mouse ES cell growth, including STAT3, are already activated in human ES cells by other factors. The recent development of feeder-independent culture conditions for human ES cells still necessitates the use of (100%) conditioned medium from feeder cells⁶⁵, indicating either that these cells require factors produced by feeder cells or that feeder cells remove some inhibitory factor from the culture medium.

The isolation of EG cells from PGCs requires LIF and, in addition, two other growth factors: Kit ligand (acting through the c-Kit receptor) and basic fibroblast growth factor (bFGF or FGF2) (acting through an FGF receptor)^{12,13}. Once established, EG cells apparently no longer require bFGF for their growth²⁸. Both human and mouse EG cells can be derived using the same combination of factors, suggesting that some of the mechanisms regulating the development of the germ line have been conserved during mammalian evolution^{12,13,15}.

of differentiated cell types formed can be varied and haphazard. When grown in suspension, pluripotent stem cells will form embryoid bodies, structures originally described in teratomas and which resemble the early pre-implantation embryo. In forming embryoid bodies, stem cell differentiation may proceed in a way related to that occurring in the embryo. Many differentiated cell types can be derived, including neurons, glia, cardiomyocytes, skeletal myocytes, adipocytes and haematopoietic cells. Differentiated cell types must be sorted out from each other and away from stem cells. This can be carried out by traditional methods of fluorescence activated cell sorting (FACS) if suitable cell-surface markers are available³³, by selective growth of differentiated cells if suitable culture conditions are known^{34–36}, or by introduction of a selectable marker that allows either FACS separation of differentiated cells or drug selection to ablate stem cells and other unwanted cells³⁷ (Fig. 3).

The basic biology of human development

Although great attention has been given to the potential use of human pluripotent stem cells in cell-based therapy, stem cells could be equally important in dissecting the development of our own species. Many of the *in vitro* techniques that have proven so useful for analysing differentiation in murine ES cells can now be applied to human cells. These include directed differentiation, gene trapping, lineage marking, cell ablation and lineage selection²⁴. In the not too distant future we may be able to build a complete gene-expression road map of how different cell types are formed, survive, proliferate, differentiate and migrate during development. This information will also help us to understand how embryogenesis can go wrong. The effects of teratogens, agents that cause fetal malformations, could be defined much more clearly than in the past and screening for new teratogens could be facilitated³⁸. Studies on the basic biology of human pluripotent stem cells could also lay the fundamental groundwork for future clinical applications.

Do we need stem cells from embryos?

Because pluripotent stem cells are able to form virtually any cell type present in the body, many new disease treatments become possible.

In theory, neurons and glia could be produced to treat neurodegenerative diseases such as Parkinson's and Alzheimer's, muscle cells could be produced to treat muscular dystrophies and heart disease, haematopoietic stem cells could be produced to treat leukaemias and AIDS. And the list goes on. But, pluripotent stem cells are derived from discarded human embryos. For some in society this fact is an insurmountable ethical problem. So why not use stem cells from adults instead?

Adult stem cells were thought originally to have a limited potential for production of differentiated derivatives. But recent studies have questioned that view. These studies show, for example, that neural stem cells can form blood-forming and muscle tissue³⁹, mesenchymal stem cells can produce differentiated cell types in the brain⁴⁰, and skin stem cells can make neurons, glia, smooth muscle and adipocytes⁴¹. Such findings have had a significant impact on the debate on the derivation and use of pluripotent stem cells derived from embryos.

The main difference between embryo-derived pluripotent stem cells and so-called multipotent stem cells from embryos or adult animals is in the number of types of differentiated cells that can be produced. This may reflect the different origins of the cells. Pluripotent stem cells are derived from germ cells or cells that can make germ cells^{4,10–13}. All other stem cells are derived from cells of the animal body or soma (so-called somatic cells) that are no longer capable of making germ cells. In adult animals it is only the germ cells that retain the ability to make a new organism, a property known as developmental totipotency^{42,43}. Although nuclear cloning experiments demonstrate that the nuclei of somatic cells can be re-programmed to allow them to become totipotent and recapitulate development^{44,45}, in the normal life cycle of animal species they cannot and do not do so. Despite several studies showing that many multipotent stem cells are capable of forming a wider variety of cell types than previously thought, it is unlikely that they can make the range of cell types made by the embryo-derived pluripotent stem cells. Consequently, adult stem cells may be immensely useful for treatment of some human diseases, but simply unable to make certain cell types required for treatment of other diseases.

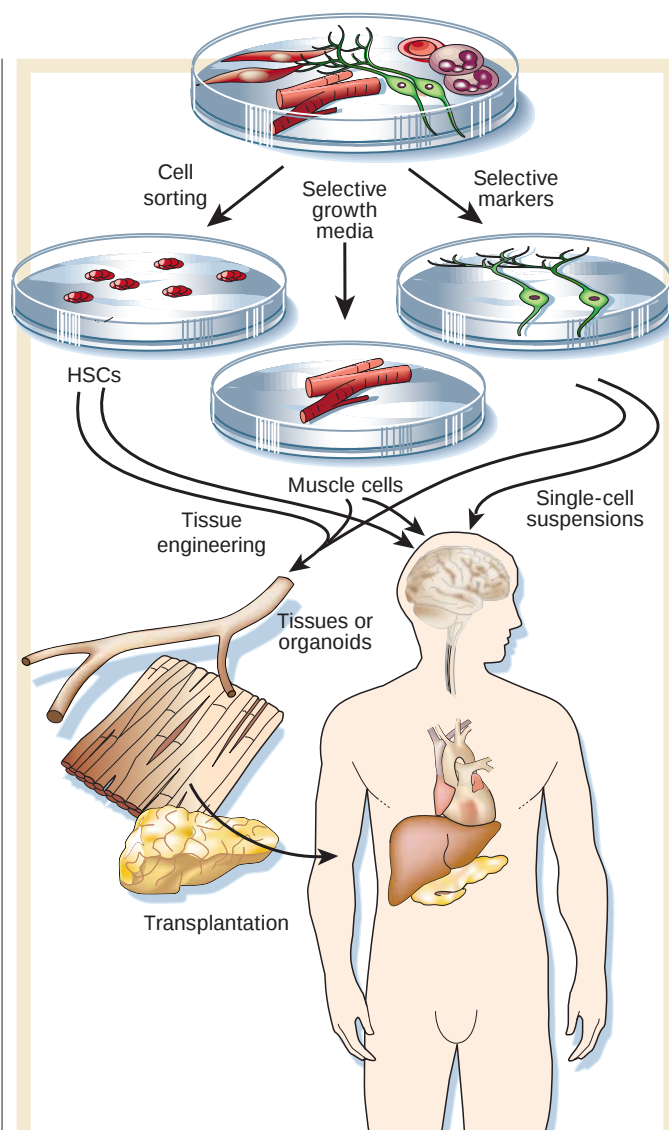


Figure 3 Isolation and separation of differentiated cells from pluripotent stem cells. Differentiation of pluripotent stem cells occurs in a haphazard manner producing many types of differentiated cell. Such differentiated cells can be isolated from other cells, including stem cells, by fluorescence-activated cell sorting (FACS), by culture in conditions that favour one cell type over another, or by the use of selectable markers such as green fluorescent protein or resistance to neomycin. These selectable markers must be expressed from a promoter construct that is introduced (by electroporation or transfection) into the starting population of pluripotent stem cells and that drives expression of the marker in the differentiated cell type of interest. Cells can be transplanted as pure populations or, following tissue engineering, as tissues or physiologically functional parts of organs (organoids).

Adult stem cells have other significant differences from pluripotent stem cells derived from embryos. These may reflect inherent differences between the two cell types or simply technical problems in growing adult stem cells. For example, *in vitro* culture conditions have been established that allow pluripotent stem cells to be expanded seemingly indefinitely without losing differentiation potential, and in which they maintain a normal chromosome number or karyotype. In addition, murine ES and EG cells can be genetically manipulated using the technique of homologous recombination, and it seems likely that this technique could soon be applied to the human cells. Similar conditions have not been established for most adult stem cells. For example, adult haematopoietic stem cells, defined as long-term repopulating cells, cannot be expanded in culture without losing developmental potential⁴⁶. But recent studies show that one type of multipotent adult stem cell, the oligodendro-

cyte precursor cell, can be grown indefinitely in culture⁴⁷. These studies suggest that it might be possible to establish conditions in which many adult multipotent stem cells can be grown indefinitely. Nevertheless, the problems currently associated with expanding, differentiating and genetically manipulating multipotent stem cells impose certain constraints upon their use. These constraints might preclude, for example, the use of patient-derived multipotent stem cells for treating inherited disorders.

Much has been written on the relative merits of stem cells obtained from embryonic, fetal and adult sources. Admittedly, it is difficult at this time to appropriately compare stem cells from these sources, as many of the claims have not appeared as peer-reviewed publications. The most candid and realistic appraisal of these sources has been published recently (11 September 2001) by the US National Academy of Sciences. Until we are able to test stem cells from various sources side by side in the laboratory in a variety of experimental paradigms, we will not know unequivocally the 'best' source of stem cells for a given therapy. Therefore, at the present time the weight of evidence suggests there are good reasons to want to continue to work with pluripotent stem cells derived from embryos and to vigorously pursue their potential for treatment of human disease. The important question is what is really possible?

Bringing stem cells to the clinic

Successes in animal model studies

So far, there have been few demonstrations that derivatives of stem cells can be transplanted successfully in animal models of diseases or injuries, but the demonstrations have been remarkable. Cardiomyocytes selected in culture from mouse ES cells could form stable, apparently functioning intracardiac grafts in mice⁴⁸. Mouse ES cell-derived glial precursors, transplanted into a rat with myelin disease, interact with the host neurons to produce myelin in the brain and spinal cord⁴⁹. Retinoic acid-treated embryoid bodies from mouse ES cells, when transplanted into a rat spinal cord nine days after traumatic injury, differentiated into astrocytes, oligodendrocytes and neurons, and promoted motor recovery⁵⁰. A genetically selected, insulin-producing cell line derived from mouse ES cells, when implanted into the spleens of streptozotocin-induced diabetic mice, result in normalized glycaemia⁵¹.

Initial results from studies using human pluripotent stem cells are promising. In the first report using human pluripotent stem cell derivatives in a transplantation paradigm, rats with a diffuse motor neuron injury showed partial recovery of motor function (D. Kerr *et al.*, submitted). Neuronal cells derived from human EC cells after treatment with retinoic acid have improved motor and cognitive deficits in rats with stroke^{52,53} and are now being used in a safety trial for patients with basal ganglia stroke⁵⁴. Although the mechanisms for improvements in the spinal cord and stroke studies remain to be determined, the results are encouraging. The transplanted cells could be substituting directly for lost populations of cells such as neurons or glia, or they could be providing factors that facilitate the regeneration of host cells.

The results of these few studies are consistent with the belief that cell-based therapies from derivatives of pluripotent stem cells may prove effective in ameliorating the effects of some devastating diseases and injuries.

Stem cell expansion and differentiation

If stem cells are to be used to treat a wide variety of human diseases, then we will need to overcome several formidable challenges. Stem cells will be needed in large quantities and be able to differentiate in a controlled manner to form homogeneous populations of cells that are histocompatible with an individual.

How can stem cells be grown in large scale? Human stem cell populations proliferate more slowly than their murine counterparts, differentiate more readily and their cloning efficiency is very low^{14,55}. Recent studies show that addition of basic fibroblast growth factor

inhibits their tendency to differentiate and at the same time improves cloning efficiency⁵⁶. But still conditions need to be improved. Efforts to grow and manipulate mouse ES cells were aided by the availability of multiple ES cell lines for which the right methods for growth and subculture could be determined. Today a limited number of human pluripotent stem cell lines exist throughout the world and only a few pluripotent stem cell lines have been described in peer-reviewed journals^{14–16,55}. The number of usable cell lines is likely to be smaller still. Some lines will be unavailable to many researchers because of problems associated with material transfer agreements and patents. Some cell lines may not survive in long-term culture and others that do survive may do so by carrying gene mutations or chromosomal alterations. Undoubtedly, the more cell lines available to work with the more we will learn about their basic biology.

Some advances in defining the optimal conditions for growing human pluripotent stem cells are likely to come from genomics. Gene-expression analysis of stem cells using microarray technology can provide key insights into the growth factors, growth-factor receptors and cell-adhesion molecules produced by stem cells^{57,58}. This type of analysis should allow the biology of pluripotent stem cells to be defined in a way that was unimaginable for the murine cells some 20 years ago. The human genome sequence has now also provided a plethora of new growth factors in which to grow pluripotent stem cells.

One approach to the problem of expansion of human pluripotent stem cells is to differentiate them into other progenitor cells that are easier to grow and expand^{16,59}. Previous studies using mouse EC and ES cells have shown that they can be induced to differentiate in culture into embryoid bodies. Similar studies have now been carried out with human pluripotent stem cells. Human embryoid bodies can be used to derive cell lines that have the hallmarks of precursor or progenitor cells, that can be expanded in culture, and that can differentiate into a wide variety of derivatives, including neural cells, vascular endothelium, muscle cells and endodermal derivatives⁵⁹. They also maintain a normal karyotype, can be cloned, cryopreserved and transfected. Differentiation of ES cells can also be induced by growth at high density. In such conditions, neural progenitors can be formed which themselves can be induced to form neurons¹⁶. Production of expandable cell populations from pluripotent stem cells overcomes some of the problems associated with growth of the stem cells themselves.

Safety considerations in cell-based therapies

What are likely to be the problems associated with the use of stem cells in the treatment of human diseases? As the gene therapy field has learned, human disease treatment must be both safe and effective. Three key safety issues are apparent. The first is whether cells can be derived that are histocompatible with every individual. Because human populations are genetically diverse, most types of transplantation have to overcome the problem of tissue rejection. Immune suppression and tolerance induction are two possible solutions, but both are short-term answers. Because stem cells are amenable to genetic manipulation, there may be better ways to address the problem.

Two methods can be used to generate embryo-derived pluripotent stem cells that are compatible with any individual. The first method requires the creation of an embryo by isolating a somatic nucleus from the patient and reprogramming it in an oocyte cytoplasm — so-called therapeutic cloning or somatic cell nuclear transfer. The embryo is then allowed to develop and stem cells are derived from it that are genetically identical to the patient. The second method requires that existing stem cell lines are genetically modified by homologous recombination to create a stem cell that is compatible with the patient. Presently both techniques are difficult and remain significant hurdles to the use of pluripotent stem cells for the treatment of disease phenotypes. Notably though, recent studies have shown that human ES cells can be transfected, and undergo

drug selection and clonal expansion (albeit at low efficiency)^{56,55}. So three of the steps required for gene replacement or modification through homologous recombination have been accomplished successfully in human ES cells.

The second safety issue is whether transplanted pluripotent stem cells will form tumours or otherwise differentiate improperly or inappropriately after transplantation. The ability of EC, ES and EG cells to form tumours in histocompatible animals reinforces the idea that it might be advantageous to use differentiated cells rather than stem cells for transplantation. Some studies suggest that imprinted loci are erased in some EG cell lines, which could adversely affect their growth characteristics^{60,61}. But other studies show that many EG cells behave normally in chimaeras^{13,26–28}. Similarly, it has been reported that murine ES cells show epigenetic instability and that animals formed entirely from ES cells rarely survive^{62,63}. But again, when ES cells are introduced into embryos in the presence of other cells (mirroring the proposed transplant situation) they differentiate into apparently normal functional cells²⁴.

Nevertheless, techniques need to be established for differentiating human pluripotent stem cells homogeneously and for separating them from stem cells, similar to those established for murine ES cells³⁶. These techniques will require the use of cell-surface markers (for FACS separation of cells) and differential growth conditions (to inhibit stem cell growth or promote differentiated cell growth), together with the introduction of selectable markers. The recent development of techniques for transfecting human ES cells will certainly facilitate the last method⁵⁵. But if cells are introduced into specific sites, how can we make sure that they migrate to the right sites and, perhaps more important, make sure that they do not go to the wrong place. What particular differentiation stage of cells should be used? These are some of the fundamental questions that must be addressed experimentally prior to bringing pluripotent stem cells to the clinic.

A third safety issue is associated with infectious agents that could be present in embryo-derived pluripotent stem cells or acquired by stem cells in feeder-dependent culture containing bovine serum. Feeder-cell-independent culture conditions have been developed for murine ES cells⁶⁴ and more recently for human ES cells⁶⁵. But the human ES cells still require either fetal calf serum or conditioned medium produced by mouse feeder cells for their growth⁶⁵. Ultimately it may be necessary to develop conditions for establishing and growing pluripotent human stem cells in defined, serum-free medium with purified recombinant growth factors and on defined extracellular matrices.

An important question that still has to be addressed is what criteria will be used by regulatory agencies (such as the Food and Drug Administration in the United States) for cell-based therapies. Will transplantation of single cell types be sufficient or will we need to create tissues or 'organoids' for treatment of some diseases. It is likely that many of these problems will be solved in non-human primate models before moving to the clinic. It may be necessary to work out grafting modalities for different cell types or for different tissue types that are specific for each disease.

The future brings formidable challenges

Only time will tell whether the results of cell transplantation in animal models can be recapitulated in humans and whether it will prove impossible to make certain cell types from pluripotent stem cells. Without doubt the use of mouse ES cells has contributed enormously to our understanding of the events of embryonic, fetal and postnatal development as well as adult homeostasis in that organism and, arguably, in mammals in general. Although the use of human pluripotent stem cells could well spell a new era in human medicine, such cells also promise to provide a unique window into the growth and development of our own species. Like murine EC and ES cells before them, human pluripotent stem cells can be used in *in vitro* assays to dissect the mechanisms that guide early development. Twenty years have passed since the isolation of mouse ES cells, an

event that revolutionized the mouse as an experimental organism. Now the availability of human pluripotent stem cells promises to revolutionize our understanding of our own development and the treatment of human disease. But there are still hurdles to be overcome if human pluripotent stem cells are to be used to treat human disease. On the road to the clinic we are not at the beginning of the end but, perhaps, at the end of the beginning. □

1. Stevens, L. C. The biology of teratomas. *Adv. Morphogen.* **6**, 1–31 (1967).
2. Jiang, L. I. & Nadeau, J. H. 129/Sv mice—a model system for studying germ cell biology and testicular cancer. *Mammal. Genome* **12**, 89–94 (2001).
3. Kleinsmith, L. J. & Pierce, G. B. Multipotentiality of single embryonal carcinoma cells. *Cancer Res.* **24**, 1544–1552 (1964).
4. Stevens, L. C. Origin of testicular teratomas from primordial germ cells in mice. *J. Natl Cancer Inst.* **38**, 549–552 (1967).
5. Skakkebaek, N. E., Berthelsen, J. G., Giwercman, A. & Muller, J. Carcinoma-in-situ of the testis: possible origin from gonocytes and precursor of all types of germ cell tumours except spermatocytoma. *Int. J. Androl.* **10**, 19–28 (1987).
6. Andrews, P. W. Teratocarcinomas and human embryology: pluripotent human EC cell lines. *Acta Pathol. Microbiol. Immunol. Scand.* **106**, 158–167; discussion 167–168 (1998).
7. Kahan, B. W. & Ephrussi, B. Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *J. Natl Cancer Inst.* **44**, 1015–1036 (1970).
8. Rosenthal, M. D., Wishnow, R. M. & Sato, G. H. In vitro growth and differentiation of clonal populations of multipotential mouse cells derived from a transplantable testicular teratocarcinoma. *J. Natl Cancer Inst.* **44**, 1001–1014 (1970).
9. Evans, M. J. The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J. Embryol. Exp. Morphol.* **28**, 163–176 (1972).
10. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
11. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
12. Resnick, J. L., Bixler, L. S., Cheng, L. & Donovan, P. J. Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550–551 (1992).
13. Matsui, Y., Zebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847 (1992).
14. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
15. Shamblo, M. J. *et al.* Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl Acad. Sci. USA* **95**, 13726–13731 (1998).
16. Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nature Biotechnol.* **18**, 399–404 (2000).
17. Thomson, J. A. *et al.* Isolation of a primate embryonic stem cell line. *Proc. Natl Acad. Sci. USA* **92**, 7844–7848 (1995).
18. Thomson, J. A. *et al.* Pluripotent cell lines derived from common marmoset (*Callithrix jacchus*) blastocysts. *Biol. Reprod.* **55**, 254–259 (1996).
19. Pera, M. F., Reubinoff, B. & Trounson, A. Human embryonic stem cells. *J. Cell Sci.* **113**, 5–10 (2000).
20. Pesce, M., Gross, M. K. & Scholer, H. R. In line with our ancestors: Oct-4 and the mammalian germ. *BioEssays* **20**, 722–732 (1998).
21. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
22. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).
23. Martin, G. R. Teratocarcinomas and mammalian embryogenesis. *Science* **209**, 768–776 (1980).
24. Smith, A. in *Stem Cell Biology 205–230* (Cold Spring Harbor Laboratory Press, 2001).
25. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255–256 (1984).
26. Labosky, P. A., Barlow, D. P. & Hogan, B. L. Embryonic germ cell lines and their derivation from mouse primordial germ cells. *Ciba Found. Symp.* **182**, 157–168; discussion 168–178 (1994).
27. Labosky, P. A., Barlow, D. P. & Hogan, B. L. Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197–3204 (1994).
28. Stewart, C. L., Gadi, I. & Bhatt, H. Stem cells from primordial germ cells can reenter the germ line. *Dev. Biol.* **161**, 626–628 (1994).
29. Brinster, R. L. The effect of cells transferred into the mouse blastocyst on subsequent development. *J. Exp. Med.* **140**, 1049–1056 (1974).
30. Illmensee, K. & Mintz, B. Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts. *Proc. Natl Acad. Sci. USA* **73**, 549–553 (1976).
31. Papaioannou, V. E., Gardner, R. L., McBurney, M. W., Babinet, C. & Evans, M. J. Participation of cultured teratocarcinoma cells in mouse embryogenesis. *J. Embryol. Exp. Morphol.* **44**, 93–104 (1978).
32. Papaioannou, V. E., McBurney, M. W., Gardner, R. L. & Evans, M. J. Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258**, 70–73 (1975).
33. Nishikawa, S. I., Nishikawa, S., Hirashima, M., Matsuyoshi, N. & Kodama, H. Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. *Development* **125**, 1747–1757 (1998).
34. Wiles, M. V. & Keller, G. Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. *Development* **111**, 259–267 (1991).
35. Nakano, T., Kodama, H. & Honjo, T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098–1101 (1994).
36. Li, M., Pevny, L., Lovell-Badge, R. & Smith, A. Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr. Biol.* **8**, 971–974 (1998).
37. Muller, M. *et al.* Selection of ventricular-like cardiomyocytes from ES cells in vitro. *FASEB J.* **14**, 2540–2548 (2000).
38. Eiges, R. *et al.* Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Curr. Biol.* **11**, 514–518 (2001).
39. Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells in vivo. *Science* **283**, 534–537 (1999).
40. Mezey, E., Chandross, K. J., Harta, G., Maki, R. A. & McKeercher, S. R. Turning blood into brain: cells bearing neuronal antigens generated in vivo from bone marrow. *Science* **290**, 1779–1782 (2000).
41. Toma, J. *et al.* Isolation of multipotent adult stem cells from the dermis of mammalian skin. *Nature Cell Biol.* **3**, 778–784 (2001).
42. McLaren, A. Mammalian germ cells: birth, sex, and immortality. *Cell Struct. Funct.* **26**, 119–122 (2001).
43. Pesce, M. & Scholer, H. R. Oct-4: control of totipotency and germline determination. *Mol. Reprod. Dev.* **55**, 452–457 (2000).
44. Campbell, K. H., McWhir, J., Ritchie, W. A. & Wilmut, I. Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**, 64–66 (1996).
45. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
46. Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
47. Tang, D. G., Tokumoto, Y. M., Apperly, J. A., Lloyd, A. C. & Raff, M. C. Lack of replicative senescence in cultured rat oligodendrocyte precursor cells. *Science* **291**, 868–871 (2001).
48. Klug, M. G., Soonpaa, M. H., Koh, G. Y. & Field, L. J. Genetically selected cardiomyocytes from differentiating embryonic stem cells form stable intracardiac grafts. *J. Clin. Invest.* **98**, 216–224 (1996).
49. Brustle, O. *et al.* Embryonic stem cell-derived glial precursors: a source of myelinating transplants. *Science* **285**, 754–756 (1999).
50. McDonald, J. W. *et al.* Transplanted embryonic stem cells survive, differentiate and promote recovery in injured rat spinal cord. *Nature Med.* **5**, 1410–1412 (1999).
51. Soria, B. *et al.* Insulin-secreting cells derived from embryonic stem cells normalize glycemia in streptozotocin-induced diabetic mice. *Diabetes* **49**, 157–162 (2000).
52. Borlongan, C. V., Tajima, Y., Trojanowski, J. Q., Lee, V. M. & Sanberg, P. R. Cerebral ischemia and CNS transplantation: differential effects of grafted fetal rat striatal cells and human neurons derived from a clonal cell line. *NeuroReport* **9**, 3703–3709 (1998).
53. Muir, J. K. *et al.* Terminally differentiated human neurons survive and integrate following transplantation into the traumatically injured rat brain. *J. Neurotrauma* **16**, 403–414 (1999).
54. Kondziolka, D. *et al.* Transplantation of cultured human neuronal cells for patients with stroke. *Neurology* **55**, 565–569 (2000).
55. Scholz, G., Pohl, L., Genschow, E., Klemm, M. & Spielmann, H. Embryotoxicity screening using embryonic stem cells in vitro: correlation to in vivo teratogenicity. *Cells Tissue Organs* **165**, 203–211 (1999).
56. Amit, M. *et al.* Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.* **227**, 271–278 (2000).
57. Phillips, R. L. *et al.* The genetic program of hematopoietic stem cells. *Science* **288**, 1635–1640 (2000).
58. Tersikh, A. V. *et al.* From hematopoiesis to neurogenesis: evidence of overlapping genetic programs. *Proc. Natl Acad. Sci. USA* **98**, 7934–7939 (2001).
59. Shamblo, M. J. *et al.* Human embryonic germ cell derivatives express a broad range of developmentally distinct markers and proliferate extensively in vitro. *Proc. Natl Acad. Sci. USA* **98**, 113–118 (2001).
60. Tada, M., Tada, T., Lefebvre, L., Barton, S. C. & Surani, M. A. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J.* **16**, 6510–6520 (1997).
61. Tada, T. *et al.* Epigenotype switching of imprintable loci in embryonic germ cells. *Dev. Genes Evol.* **207**, 551–561 (1998).
62. Humphreys, D. *et al.* Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95–97 (2001).
63. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
64. Nichols, J., Evans, E. P. & Smith, A. G. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341–1348 (1990).
65. Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnol.* **19**, 971–974 (2001).
66. Martin, G. R. & Evans, M. J. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc. Natl Acad. Sci. USA* **72**, 1441–1445 (1975).
67. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).
68. Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687 (1988).
69. Yoshida, K. *et al.* Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech. Dev.* **45**, 163–171 (1994).
70. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048–2060 (1998).
71. Matsuda, T. *et al.* STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J.* **18**, 4261–4269 (1999).
72. Burdon, T., Tracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210**, 30–43 (1999).

Acknowledgements

We apologize to the colleagues whose work was not cited because of space constraints. We are indebted to L. F. Lock for critical comments on the manuscript and with L. Cheng, D. Panchison, H. Scholer, M. Bartolomei, L. Iacovitti and J. McLaughlin for helpful discussions. We thank M. Linkinhoker for help with the figures.

Stem cells find their niche

Allan Spradling, Daniela Drummond-Barbosa & Toshie Kai

HHMI/Department of Embryology, Carnegie Institution of Washington, Baltimore, Maryland 21210, USA

The concept that stem cells are controlled by particular microenvironments known as 'niches' has been widely invoked. But niches have remained largely a theoretical construct because of the difficulty of identifying and manipulating individual stem cells and their surroundings. Technical advances now make it possible to characterize small zones that maintain and control stem cell activity in several organs, including gonads, skin and gut. These studies are beginning to unify our understanding of stem cell regulation at the cellular and molecular levels, and promise to advance efforts to use stem cells therapeutically.

Stem cells fascinate us from both a theoretical and a practical viewpoint. Their defining characteristics — extensive proliferative potential and an ability to give rise to one or more differentiated cell types — are common in early mammalian embryos. But embryonic cells lose these properties as differentiation ensues and growth-promoting signals decline. By adulthood, the few remaining stem cells are dispersed and virtually invisible; however, the surviving adult stem cells have achieved something remarkable. They can operate at 'steady state' — that is, they can generate on average one replacement stem cell and one tissue cell at each division with no apparent limit.

The mechanisms that stem cells use to accomplish self-renewal promise fundamental insight into the origin and design of multicellular organisms. These pathways make it possible to repair damage and extend organismal life beyond that of component cells, and probably preceded the evolution of complex metazoans. Can knowledge of stem cells be used to develop a cell-based medicine to repair malformed, damaged or ageing tissues? We are drawn to the subject of stem cells as we are to stories of perpetual motion machines or the fountain of youth; understanding the inner workings of stem cells might help to realize their mythic promise.

The true nature of stem cells can be learned only by discovering how they are regulated. One possibility is that internal mechanisms in the stem cell control a stereotypic pattern of unequal division that generates daughters with different fates. Such autonomously acting stem cells would probably exhibit specialized, tissue-specific patterns of gene expression. It is becoming increasingly clear, however, that the specialized functions required to ensure proper stem cell function are vested in neighbouring differentiated cells. By signals and other intercellular interactions, these cells control the behaviour of adjacent stem cells that may themselves be relatively unspecialized. Consequently, location rather than special patterns of gene expression may characterize stem cells.

The realization that stable, custom microenvironments might control haematopoietic stem cells led Schofield¹ to call such regions 'niches'. For our purposes, a niche (Fig. 1) is considered to be a subset of tissue cells and extracellular substrates that can indefinitely house one or more stem cells and control their self-renewal and progeny production *in vivo*. Here we review our current understanding of stem cell niches, focusing on those systems that are best characterized at the cellular and mechanistic levels. Haematopoietic and neural stem cells are considered separately in the respective

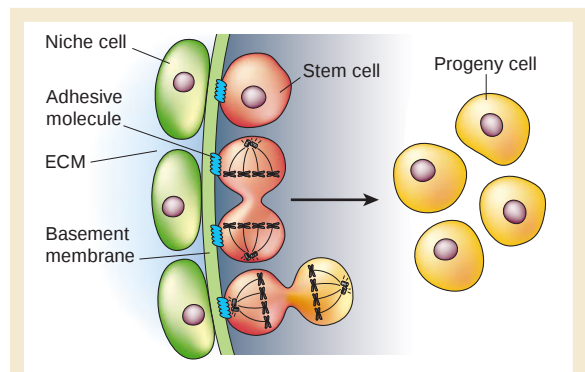


Figure 1 Niche structure. Niche cells (green) underlying a basement membrane signal to stem cells (red) to block differentiation and regulate division. When a lineage mechanism prevails (lower mitotic cell), the stem cell divides such that one daughter retains its connections to the niche, while the other (yellow) becomes untethered and begins to differentiate. When a population mechanism prevails (upper mitotic cell), stem cell division may be either symmetric (shown) or asymmetric (not shown), as determined by local factors. ECM, extracellular matrix.

reviews by Weissman and colleagues (pages 105–111) and Temple (pages 112–117) in this issue.

Locating and classifying niches

Identifying niches and determining how they are regulated is experimentally challenging. To look systematically for niches in a target tissue, stem cells first need to be identified by marking cells and following their lineages. When individual stem cells have been localized, evidence that they reside in a common microenvironment is sought by characterizing cellular neighbours, expression patterns of signalling molecules and local environmental factors such as extracellular matrices. If commonalities are found, then the microenvironment can be perturbed to learn which aspects, if any, affect stem cell behaviour.

Showing that a niche is present requires ultimately that stem cells be removed and added back (Fig. 2a). A niche should persist in the absence of resident stem cells and support the stem cell activity of competent exogenous cells. By contrast, the fate of the removed stem cells is irrelevant to the niche issue. Sometimes exogenous cells can be added directly to intact tissues and then function as stem cells, presumably by occupying empty, partially filled or ectopic niches, or by displacing endogenous stem cells. But endogenous stem cells may have to be depleted before a niche accepts donor cells. The limitations of this reconstitution

assay include the difficulty of interpreting negative results, and the uncertain biological significance of niches defined by reconstitution when information on the tissue's normal stem cells is lacking.

Niches are thought to operate in one of two basic ways. Lineage niches specify precisely individual stem cell divisions and daughter cell fates. For example, a stem cell in such a niche might contact the stroma asymmetrically and orientate its division plane to ensure that only one daughter inherits adhesive contacts with the basement membrane (Fig. 1, bottom). The daughter inheriting these contacts is held in the niche, where it will continue to be maintained as a stem cell, whereas the other daughter becomes untethered and relocates away from the stromal cells and their signals. By contrast, when a population mechanism operates (Fig. 1, middle cells), both daughters may remain as stem cells or both may differentiate. These two mechanisms can be distinguished by lineage studies in which single labelled stem cells and their progeny are followed (Fig. 2b): the number of labelled stem cells will remain unchanged by a lineage mechanism, but a population mechanism will eventually homogenize the labelling of all the niche's stem cells.

Germline niches

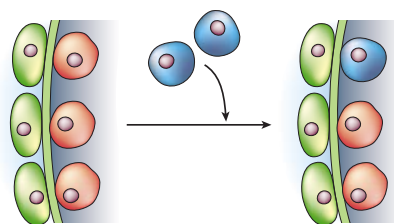
Mammalian testis

The strongest evidence for niche-based regulation in any mammalian tissue probably comes from studies of spermatogenesis. In the testis, germ cell development is maintained by germline stem cells (GSCs) known as A_{single} (A_s) spermatogonia that lie in contact with the basement membrane of the seminiferous tubule (Fig. 3a). After stem cell division, a large collection of A- and B-type spermatogonia joined by intercellular bridges forms through a series of roughly ten synchronous, incomplete divisions. The spermatogonia can move laterally along the basement membrane and are encased by Sertoli cells that mediate many aspects of male germ cell development. As germ cells enter meiosis and begin to differentiate into sperm, they lose contact with the basement membrane and move towards the lumen of the seminiferous tubule, while maintaining their association with Sertoli cells^{2,3}.

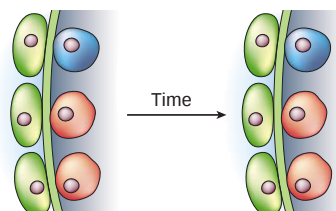
The existence of niches in the testis was demonstrated directly when early germ cells were transplanted into the seminiferous tubules of host males whose GSCs had been depleted⁴. Introduced cells colonize sites on the wall of the tubule and found growing clones of developing sperm that can restore host fertility. Each colony seems to be established by a single GSC that proliferates and differentiates in a regular manner, as it expands in both directions along a segment of the host tubule⁵⁻⁷. Most of the founder cell's initial progeny become stem cells. These cells probably occupy empty niches that lie in abundance along the depleted basal layer. Only when the colony has expanded greatly are differentiating sperm seen near the centre of the colony. Thus, the number of colonies represents a minimum estimate of the number of niches, which must in reality be much larger. Although a reconstitution assay has long been available for haematopoietic stem cell niches⁸, the simple morphology and accessibility of the testis offer experimental advantages for studying the cellular and molecular regulation of a normal niche *in vivo*. For example, the reconstituted stem cells and their surroundings can be compared with the situation in the normal testis — information that has been lacking for haematopoietic stem cells.

The reconstitution assay not only reveals the presence of thousands of germline stem cell niches along the wall of the seminiferous tubules, but also provides information on their developmental properties. For example, niche number and quality seem to change during the development of mouse testes⁹. Testes from W^+ mice lack normal c-Kit function and are deficient in endogenous spermatogenesis. As hosts, they support 8–10 times more colonies when derived from pups than from adults, and the colonies are larger in pups. This suggests that niches may be more abundant or more accessible in young than in adult seminiferous tubules. A caveat is that there may be developmental differences in the number of unoccupied niches in

a Replacement assay



b Lineage mechanism



Population mechanism

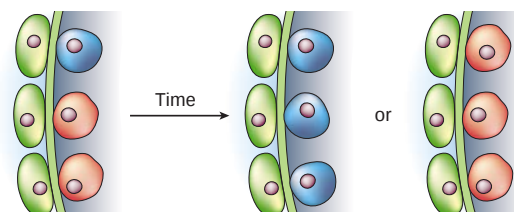


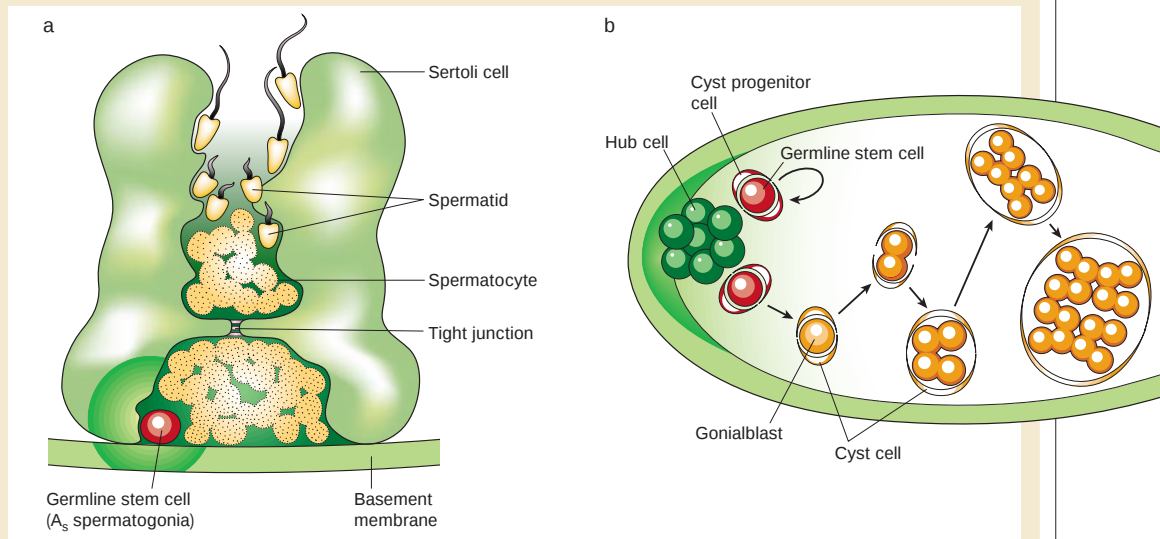
Figure 2 Manipulating niches. **a**, Replacement assay to identify niches. Labelled stem cells (blue) are introduced into a tissue. If one or more of the introduced cells becomes associated stably with a particular tissue region and subsequently functions as a stem cell, then the existence and location of a stem cell niche (green) is inferred. **b**, Cell marking experiments. Marking a single stem cell *in vivo* (blue), and following the number and location of marked progeny cells reveals its regulatory behaviour. When the stem cell follows a lineage mechanism, the number of labelled stem cells remains relatively constant over many stem cell cycles. When the initially labelled stem cell follows a population mechanism (bottom), initially mosaic niches will sort out stochastically into niches that have all labelled or all unlabelled cells.

W^- testes, because the assay depends on depleting endogenous niches¹⁰. Other studies show that compared with adult spermatogonial cells, many fewer colonies per input cell are obtained using prospermatogonia — cells that have not yet become associated with the basement membrane or Sertoli cells — from newborn mice. Thus, germ cells seem to become more competent to populate a niche as a result of developmental events that correlate with the establishment of niches *in vivo*.

Several candidate niche components and signals have been analysed. The $\alpha_6\beta_1$ integrin is enriched in the basal layer and may help to attach spermatogonial stem cells to laminin in the basement membrane¹¹. Consistent with this idea, transplant assays show that crude cell populations can be enriched for male GSCs by binding to laminin, the target of $\alpha_6\beta_1$, but not to collagen IV or fibronectin. Bone morphogenetic protein (BMP)-4 produced by the extraembryonic ectoderm stimulates the growth and development of early germ cells¹². BMP-8a and BMP-8b also are needed for germ cell development¹³, but an ongoing function for BMPs in germline stem cells has not been established as yet¹⁴. Sertoli cells produce another transforming growth factor (TGF)- β -related factor, GDNF, that affects the proliferation of pre-meiotic germline cells, including perhaps stem cells¹⁵.

Several other candidate signals seem to act at a slightly later stage. Signals mediated by SCF, a Sertoli cell-product encoded by the *Steel*

Figure 3 Male germ cell niches. **a**, A cross-section of part of a seminiferous tubule of a mouse testis is shown to illustrate the approximate location and properties of the mammalian germline stem cell niche. The stem cells also known as A_{single} (A_s) cells (red) constitute a minority of the basal germ cells in contact with the basement membrane (green). Most other basal cells (yellow) are part of growing germ cell cysts, whose members are interconnected by ring canals. As development proceeds, the cysts leave the basement membrane and move in a highly ordered manner towards the lumen of the tubule (top). All developing germ cells are surrounded by Sertoli cells (green). It is not known whether the region of the Sertoli cell or basement membrane near the stem cells are specialized (dark green). **b**, A sagittal section of the distal tip of the *Drosophila* testis is drawn schematically and leaves out most of the cells for clarity. The hub cells (green) are attached to 5–9 groups comprising one germline stem cell and two cyst progenitor stem cells (red). The basement membrane (green) may be thinner and specialized (dark green) near the hub. Germline stem cell daughters called gonialblasts are encased by two cyst cells, begin to differentiate, and move posteriorly away from the hub. After four more divisions, the cyst cells (yellow) cease division and synchronously enter meiosis.



(*Sl*) locus, are received by the germ cell c-Kit receptor and required for germ cells to develop beyond type A spermatogonial stages. Wild-type germ cells transplanted to *Sl* testes can still form colonies¹⁰, but they fail to differentiate, which places the requirement for SCF downstream from the stem cell stage. The RNA helicase Vasa is expressed in the developing germ cells of many different species, including mice. In the absence of Vasa, male germ cell proliferation is reduced severely in spermatogonial stages¹⁶. The timing of the defect argues against a requirement for Vasa to maintain germline stem cells in the niche.

Further studies at the level of individual stem cells are needed to increase our understanding of testis niches and their regulation. Do endogenous A_s spermatogonia reside at special positions along the wall of the seminiferous tubule, or does the entire membrane, in collaboration with local Sertoli cells, constitute a stem cell niche? If there are special sites, how are they distributed, and do transplanted A_s cells have to occupy them, or can they colonize 'new' niches not used by normal stem cells? Do male germline stem cells maintain themselves using a population mechanism, or do the two daughter cells normally acquire different fates?

Drosophila testis

Germline stem cells in model invertebrates are located near the blind end of simple tubes. The ends of both male and female *Drosophila* gonads contain germline stem cells, whose behaviour has been verified by cell marking^{17,18} and which have been proposed to reside in niches^{19–21}. In the adult testis (Fig. 3b), a cluster of 10–12 cells form the 'hub' distally, in contact with the basement membrane. About 5–9 GSCs and 10–18 cyst progenitor stem cells (somatic stem cells or SSCs) are interspersed around the hub in a ring, with part of each cell directly contacting the hub.

Both GSCs and SSCs divide radially outwards from the hub. As expected for a lineage niche, one GSC daughter remains in the same environment adjacent to the hub and continues as a stem cell, whereas the other enters a different environment, acquires new cell contacts, and differentiates as a gonialblast. In larval testes, both the number of hub cells and the number of associated stem cells are twice as large as in young adults²². In older adults, the number of these cells decreases further²¹. This decrease may parallel the developmental decline in the ability of the mammalian testis to support the growth of

transplanted stem cells⁹. But proof that the *Drosophila* testis functions as a stem cell niche will require reconstituting stem cell function using exogenous cells.

Signalling by JAK–STAT (Janus kinase/signal transducer and activator of transcription) has been shown to be important in controlling male GSCs (E. Matunis, personal communication). The *unpaired* (*upd*) gene encodes a putative ligand that activates this pathway through an unidentified receptor. Upd is produced by hub cells and can stimulate stem cells when expressed ectopically in the testis.

Other intercellular signals are also likely to regulate stem cells and their progeny. When newly formed somatic cyst cells contain somatic clones of *raf*²¹, differentiation of nascent gonialblast and germ cell cysts is slowed, causing these early stages to accumulate at the testis tip. Under these conditions, associations between somatic cyst cells and germ cells are abnormal, and the number of GSCs associated with the hub does not decline as fast as in wild-type males of the same age. Reductions in epidermal growth factor receptor (*Egfr*) activity produce similar effects through an action on somatic cells³. This shows that a somatic *Egfr* and Raf-mediated signal is important for gonialblast and spermatogonial development, but it is less clear whether the signals or the disruptions in germ cell development are responsible for the reduced turnover of stem cells.

Drosophila ovariole

The tip of each ovariole contains specialized cells that contribute to its ability to act as a niche (Fig. 4a). Two germline stem cells associate closely with a stack of about ten terminal filament cells. Like the dermal papilla of a mammalian hair follicle (see below), the terminal filament first contributes to ovariole development²³ and then supports stem cell activity^{24,25}. Five or six cap cells located asymmetrically at the base of the terminal filament directly contact the GSCs, forming junctions that associate preferentially with the stem cell's specialized organelle known as the fusome²⁶. GSCs divide along the anterior–posterior axis²⁷, which ensures that the anterior germ cell daughter inherits a cap cell contact and the stem cell fusome, and remains a stem cell. The posterior daughter is blocked from cap cell access, and is drawn away from the ovariole tip where it differentiates into a cystoblast. Although GSCs divide asymmetrically with respect to the fusome, this inequality can be eliminated by mutating a fusome structural component without affecting stem cell activity²⁸. Thus,

GSCs utilize a lineage mechanism, but their unequal partitioning of cytoplasmic components is either unimportant or redundant for stem cell function.

To prove that ovarioles contain a stem cell niche, one would ideally remove the existing stem cells, reintroduce exogenous stem cells, and show they can re-occupy and come under the control of the remaining stromal cells. 'Empty niches' can be created genetically by eliminating maternally required genes such as *oskar* or *tudor*, but a method for introducing exogenous cells is not available. Nonetheless, the presence of a niche has been shown directly by a different approach²⁹. One GSC was marked and genetically destabilized so that it would begin to differentiate and move away. Ovarioles were then identified in which the marked GSC had recently differentiated and opened-up space adjacent to the cap cells. In all cases, a new germ cell was found to arise by symmetric division of an adjacent stem cell, and to fill such a vacancy almost immediately. Replacement also occurs at a low level in wild-type adults. These observations suggest that conditions inside a niche, such as the availability of a cap cell-binding site, influence whether a lineage-based or population-based mechanism is used.

The BMP homologue encoded by *decapentaplegic* (*dpp*) functions as a chief signal for female GSCs^{19,29}. Loss of Dpp reception causes stem cell differentiation, whereas excess Dpp blocks differentiation and leads to stem cell tumours. Several somatic cell types near the stem cells express *dpp* messenger RNA, but it remains unclear whether the final active signal is distributed locally or is widespread. Stem cells are the only germ cells located near the niche that require *dpp* signalling, but this pathway is involved in several additional processes later in gametogenesis^{19,30}.

Several other genes have been identified that may contribute to operating the ovariole niche. The *piwi* gene encodes a phylogenetically conserved nucleoplasmic protein that is expressed in both somatic niche cells and germ cells³¹. Somatically, *piwi* is needed to maintain stem cells, while loss of *piwi* in GSCs slows their division and disrupts germ cell development; the production of male germ cells is also greatly reduced and males are sterile. Genes related to *piwi* are known or likely to function in a wide variety of stem cells, making it one of the best current candidates for a 'stem cell' gene³²⁻³⁴. The *fs(1)Yb* gene is also needed for continuing female GSC activity, and may act primarily to maintain *piwi* expression in terminal filament and cap cells²⁵. The putative translational regulators Pumilio and Nanos are specifically required to maintain female GSCs^{28,35}. The activity of these genes may be modulated in response to niche signals.

Developing male and female germ cells both require the *bag-of-marbles* (*bam*) and *benign gonial cell neoplasm* (*bgn*) genes^{36,37}, but whereas these genes function in cystoblasts in females, they are needed to terminate cyst growth in males. Expressing Bam in female GSCs causes them to differentiate, but the same treatment does not perturb male GSCs. The striking differences between male and female GSC may reflect their different evolutionary origins. The male niche may be evolutionarily old, as active stem cells are found in adults from a wide range of species. In contrast, most female gonads, even among insects, do not retain stem cells in adulthood. The female stem cell niche might have arisen relatively recently during Dipteran evolution, perhaps by maintaining into adulthood mechanisms that stimulate germ cell growth in the embryo.

Caenorhabditis elegans

In the hermaphroditic gonad of *C. elegans* adults, a cluster of actively cycling syncytial cells at the distal end maintains the production of germ cells (Fig. 4b). A somatic distal-tip cell that expresses the Lag-2 ligand³⁸ extends processes that contact these cells and stimulates mitotic division by interacting with the Glp-1 receptor³⁹⁻⁴¹. Asymmetric stem cell divisions have not been observed⁴¹, and the worm gonad represents a strong candidate for a niche that is controlled by a population mechanism. But appropriate lineage studies to

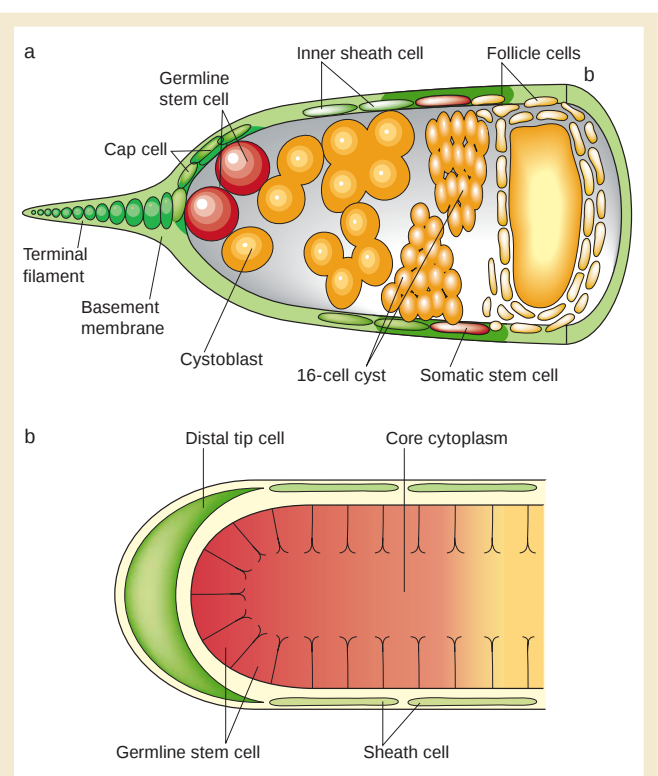


Figure 4 Female and hermaphrodite gonads. **a**, Sagittal section of a *Drosophila* germarium, showing the location of germline stem cells (red cells on left) and somatic stem cells (red cells on top and bottom). The terminal filament and cap cells (green) that make up the germ cell niche lie against a basement membrane. The cystoblast and developing cysts (yellow) move away from the anterior tip while contacting inner sheath cells (green). At the posterior zone of inner sheath cells lie 2–3 somatic stem cells (SSCs; red). SSCs contact the basement membrane (green), which may be locally specialized (dark green). The progeny of the SSCs (yellow) divide and then migrate inwards to surround the germline cysts where they become follicle cells. The SSCs receive signals from the terminal filament and cap cells that control stem cell number and proliferation rate. **b**, The distal end of the *C. elegans* gonad. The distal tip cell (green) contacts syncytial germline stem cells (red) in the mitotic zone near the distal end of the gonad. More proximally, germ cells (yellow) enter meiosis and eventually bud-off as growing oocytes.

unambiguously distinguish these mechanisms have not been carried out, nor have stem cell replacement experiments been reported.

Epithelial niches

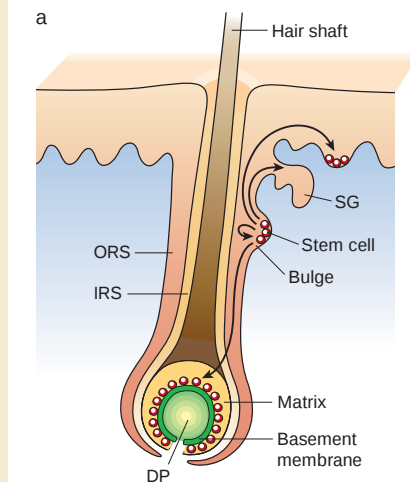
Mammalian skin

Many different stem cells maintain mammalian skin and its associated structures, such as hair follicles, sebaceous glands and sweat glands (Fig. 5a). Away from the follicles, a steady flow of differentiating keratinocytes is supplied by relatively rare epidermal stem cells located in the basal layer^{42,43} and in the deep rete ridges⁴⁴. Hair follicles contain two zones of stem-like cells. The hair shaft and its surrounding sheaths are produced by matrix cells located on a basement membrane overlying the protruding piece of dermis known as the dermal papilla. Longer-lived stem cells reside more apically within the 'bulge' region^{45,46}. Cell marking and follicle transplantation experiments have shown that the bulge contains stem cells that repopulate the matrix region, form new sebaceous glands, and replace epidermal stem cells^{47,48}.

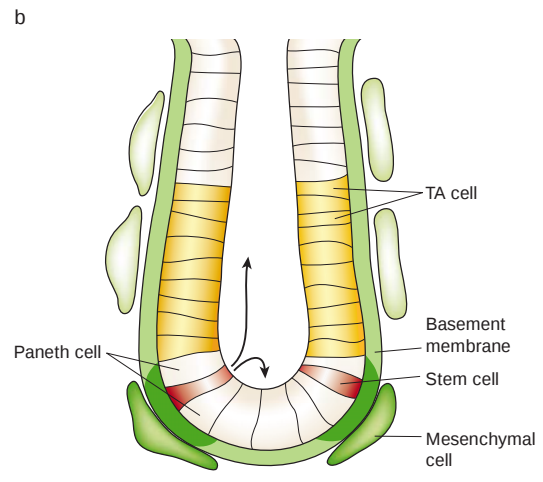
A stem cell replacement assay is not available for mammalian epidermis to test directly whether niches exist. Because bulge stem cells have not been mapped with single-cell resolution, they have yielded few clues concerning their mode of regulation. However, the normal developmental cycle of hair growth (anagen) and degrowth (catagen)⁴⁹

Figure 5 Epithelial stem cell niches.

a. Mammalian epidermal stem cells. A hair follicle and a segment of adjacent skin is illustrated. The dermal papilla (DP) signals to matrix stem cells (red) located across a basement membrane (green). Matrix cell daughters (yellow) differentiate into a variety of cell types, including the medulla, cortex and cuticle of the hair shaft (brown), the inner root sheath (IRS) and the outer root sheath (ORS). About two-thirds of the way up an anagen follicle lies the bulge — an expanded region that contains long-term stem cells (red). These cells periodically replenish (arrows) the matrix cells, and also help maintain the sebaceous gland (SG) and the epidermal stem cells (red, top layer) that lie against the basement



membrane (not shown) overlying the basal layer in interfollicular regions. **b.** A mammalian gut crypt is a tube of cells arrayed on a basement membrane (green). Stem cells (red) are located in the basal region along with Paneth cells, but their exact location is variable and both types account for only a fraction of the cells present in the regions shown. A portion of the basement membrane in the stem cell region may be specialized (dark green). Stem cell progeny (yellow) known as transit amplifying cells (TA) move upwards and differentiate. Underlying mesenchymal cells (green) send signals that help regulate stem cell activity.



provides a natural replacement assay, and leads us to regard the matrix region of the follicle as a stem cell niche. During catagen, the lower two-thirds of the follicle gradually disappears and the dermal papilla reaches the level of the bulge. When the next round of anagen begins, daughter cells derived from the bulge stem cells move onto the dermal papilla, become new matrix cells and re-initiate a hair shaft. Other bulge cell derivatives move upwards to maintain the sebaceous gland, and probably replenish surrounding epidermal stem cells.

The dermal papilla seems to be a key niche component and a source of signals that stimulate the activity of matrix stem cells. Hair follicles do not develop, persist or function without a dermal papilla. Fibroblast growth factor (FGF)-7 is produced in the dermal papilla and stimulates keratinocyte growth⁵⁰. Targeted disruption of $\beta 1$ -integrin in matrix and outer root sheath cells retards matrix cell proliferation and hair growth without grossly disrupting matrix cell attachment to the dermal papilla during the first hair cycle^{51,52}. Disruption of α -catenin leads to hyperproliferation⁵³. BMP-4 is also expressed in dermal papilla cells, and mice expressing the BMP inhibitor Noggin in matrix cells and early hair precursors have altered proliferation of hair shaft precursors⁵⁴.

Additional studies are uncovering the roles played by two other signalling pathways. Sonic hedgehog (Shh) is required in developing hair follicles⁵⁵, and like Krox-20 and TGF β RII, it is heavily expressed in the matrix along only one side of the hair bulb⁵⁶. Activating mutations in *shh* or *ptc1* frequently cause basal cell carcinomas in humans and mice, but the cellular origin of these carcinomas remains in dispute and it is unclear whether Shh normally stimulates matrix cell proliferation in its niche.

Genetic studies indicate that Wnt signalling is also required to form hair follicles^{57,58} and for matrix-derived cells to differentiate into follicular rather than epidermal keratinocytes⁵⁹. Overactivation of Wnt signalling causes mice to develop more hair follicles⁶⁰, and to yield skin- and hair-derived tumours⁶¹. Wnt3a and Wnt5a probably mediate a reciprocal maintenance signal from the follicle's epidermal components to the dermal papilla⁶². Stimulation of bulge stem cells, which express Tcf3, may explain some of these observations⁵⁸.

Drosophila follicle cells

The SSCs in the *Drosophila* ovariole seem to reside in a relatively simple niche (Fig. 4a). Two or three SSCs lie against the basement

membrane at the posterior edge of a monolayer of inner sheath cells, which they resemble both morphologically and in their patterns of gene expression. Lineage-marking experiments¹⁷ verify that the SSCs are stem cells and that their progeny give rise to at least ten different subtypes of ovarian follicle cells. The SSC division rate is coordinated with that of downstream cells and is regulated by nutrition over a fourfold range⁶³. The main evidence for an SSC niche is a low level of natural stem cell replacement^{17,64}, similar to what occurs in the germ cell niche. It has not been shown, however, that a replacement stem cell occupies the same position as the lost SSC.

SSC division is strongly influenced by the activity of the Hedgehog (Hh) pathway^{64–67}. Hedgehog is highly expressed in the terminal filament and cap cells that lie a short distance from the SSCs. Mutations in several genes that downregulate the Hh pathway decrease somatic cell production, whereas upregulation of the Hh signalling by various means causes a two- to threefold overproduction of follicle cell precursors^{64–67}. Unlike nutritional changes, altering Hh signalling affects cell output by changing the number of stem cells⁶⁴. Some of the extra SSCs are located very close to existing stem cells, suggesting that an individual niche has expanded in size to accommodate two rather than one SSC. In other cases, however, the stem cells are well separated, suggesting that new sites may also be colonized. New niches may be generated because excess Hh signalling causes a large increase in the number of cells in the stem cell region. It is not clear whether the extra stem cells and niches are as stable as normal SSCs in the endogenous niche, however, as they were lost relatively quickly⁶⁴.

Endodermal niches

Gut crypts

The final candidate stem cell niche that we consider here maintains a steady flow of cells from crypts (Fig. 5b) in the vertebrate gut to outwardly projecting villi. Crypts are small inpocketings of the gut surface lined by a single layer of columnar cells that lie against a basement membrane. Lineage-tracing experiments identify perhaps 4–5 stem cells that reside near the bottom of each crypt⁶⁸. Stem cell progeny differentiate into four main cell types — Paneth, enteroendocrine, goblet and columnar — as they migrate up the walls of the crypt. Lineage labelling, including the use of a gut-directed promoter that drives *cre* recombinase and the use of a floxed marker gene⁶⁹, has still not pinpointed the exact number and location of stem cells in the

basal region of the crypt. Mosaic crypts generated in such experiments, or by somatic mutation *in vivo*, progress with time to become either all labelled or all unlabelled^{69,70}. Such rapid equilibration is the expected result when a niche is maintained by a population mechanism rather than by asymmetric stem cell divisions.

Several specific cell types have been ablated to determine whether they contribute to the maintenance of stem cells. Between 30 and 50 Paneth cells reside near the stem cells in the basal region of the crypt, where they secrete antimicrobial 'cryptidin' peptides. When Paneth cells are ablated using a cryptidin2–diphtheria toxin construct there is no change in cell proliferation, despite the fact that these cells also produce potential growth modulators⁷¹. Targeted ablation of the secretin-producing cells in the crypt⁷² also has no effect on proliferation. Immature cells in these lineages are probably not removed in such experiments and cannot be ruled out as potential growth regulators.

Signals emanating from mesenchymal cells underlying the basal region of the crypt are thought to maintain the activity of crypt stem cells. One of the most important signals is mediated by the Wnt pathway acting through the Tcf-4 transcription factor. Stem cells are missing from the intestinal epithelium in mice lacking Tcf-4, indicating that Tcf-4-mediated signals are needed to form and maintain crypts⁷³. Conversely, mutations in Apc or β -catenin can cause constitutive nuclear Tcf-4/ β -catenin complexes, and accelerate proliferation. Disrupting the *Fkh6* gene, which encodes a transcription factor that is expressed in mesenchyme, causes intestinal cells to overproliferate⁷⁴. It is important to note, however, that because gut stem cells cannot be identified individually, it has not been possible to prove that these genes act on stem cells rather than on their proliferating progeny.

Regulation in development

Niches modulate their numbers during development and in response to environmental factors. Normally, the number of hair follicles does not increase after a mouse is born⁷⁵. In many tissues, however, niches multiply to keep pace with growth from juvenile to adult stages: there are many more crypts in adult guts than in newborns. Crypts usually duplicate by branching⁷⁶ — possibly when the number of stem cells or key stromal cells surpasses a critical threshold⁷⁷. Experimental manipulations show that certain signals also have the capacity to induce *de novo* production of niches. For example, new hair follicles form in adult skin when Wnt signalling is upregulated⁶⁰. In the *Drosophila* ovary, excess Hh signalling expands the size of the SSC niche and may create new niches⁶⁴. Perhaps the regulated expression of these signals controls niche growth during development.

Niches also modify their regulatory properties in response to changing conditions to ensure that stem cell activity parallels the organism's needs for particular differentiated cell types. Bulge stem cells proliferate in response to local wounding⁴⁷. After irradiation of the testis, the remaining stem cells are induced to divide equally with high probability to regenerate the population of stem cells and spermatogonia⁷⁸. Hormonal cycles regulate the proliferation of stem cells in breast and uterine epithelium, whereas ovariolar stem cells respond to nutritional levels⁶³.

Developmental mechanisms that activate or silence niches are just beginning to be studied at the molecular level. The hair cycle provides one of the best-studied systems. Outer root sheath cells produce FGF-5, which signals to terminate anagen⁷⁹. Consequently, mutants in the *angora* locus encoding FGF-5 have unusually long hair. Signalling through the EGF receptor seems to be required for entry into catagen. Expression of a dominant-negative EGF receptor in epidermal cells leads to long hairs⁸⁰, whereas loss of EGFR⁸¹ or TGF- α ⁸² produces wavy hair and disorientated hair follicles. The *hairless* gene encodes a transcription factor that is required to maintain follicle integrity during catagen⁸³.

Common properties

Although niches seem to be structurally and functionally diverse, there may be a few common, organizing principles. Many niches use

one or more specialized cell groups, such as the hub, the dermal papilla or the terminal filament/cap cells. A principal signal received by the stem cells in the niche often controls their behaviour, and may originate from within the specialized cells. Unpaired or Dpp fulfils this role for *Drosophila* male or female germ cells. Hedgehog may have this function in mouse and *Drosophila* epithelial cells. Overproduction of this signal can cause stem cell hyperplasia. In all cases, the actual signalling milieu in the niche is complex and involves additional components. A basement membrane may be part of most of the niches discussed. Extracellular matrices may help to structure niches spatially and modulate the concentration of adhesive and signalling molecules locally. In contrast, internally programmed mechanisms of generating asymmetry have not been found to play a significant role as yet.

Underlying similarities in niche regulation may ultimately minimize the distinction between lineage- and population-based regulation. It is becoming clear that the same niche may use either mechanism under particular circumstances. *Drosophila* GSCs normally use a lineage mechanism, but when replacing a missing stem cell a *Drosophila* female GSC divides equally and shifts its division plane so that both daughters contact cap cells. Similarly, stem cells in the mammalian testis are likely to alternate between the two mechanisms, although it is not known whether shifts in divisional orientation are involved. Clearly, identical stem cells might give rise to either mechanism depending on whether extra maintenance sites in the niche were available at the time and place that a stem cell divides. But not all niches may be this flexible. So far, there is no indication that the nematode gonad or the gut crypt operates other than by the population mode.

Implications

Stem cells may often be regulated in niches because of their developmental history. Like early embryonic cells, stem cells must have the capacity to grow and differentiate along several pathways in response to external signals. Rather than evolve complex, specialized internal mechanisms and signals to confer these characteristics, stem cells might have first appeared when niches acquired an ability to sequester, preserve and control undifferentiated embryonic cells that already had the necessary cellular properties. Such an origin would explain why most of the characterized niche regulatory mechanisms and signals also function in developing embryos. Although there are probably small differences among stem cells, this model predicts that the expression profiles of embryonic stem cells and those of stem cells from diverse tissues will be highly similar, encouraging attempts to reprogramme and interconvert them. Understanding niches and their signals will suggest excellent starting points for such efforts. □

1. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
2. de Rooij, D. G. & Grootegeed, J. A. Spermatogonial stem cells. *Curr. Opin. Cell Biol.* **10**, 694–701 (1998).
3. Kiger, A. A. & Fuller, M. T. in *Stem Cell Biology* (eds Marshak, D. R., Gardner, R. L. & Gottlieb, D.) 149–187 (Cold Spring Harbor Press, Cold Spring Harbor, 2001).
4. Brinster, R. L. & Zimmermann, J. W. Spermatogenesis following male germ-cell transplantation. *Proc. Natl Acad. Sci. USA* **91**, 11298–11302 (1994).
5. Parreira, G. G. *et al.* Development of germ cell transplants in mice. *Biol. Reprod.* **59**, 1360–1370 (1998).
6. Dobrinski, I., Ogawa, T., Avarbock, M. R. & Brinster, R. L. Computer assisted image analysis to assess colonization of recipient seminiferous tubules by spermatogonial stem cells from transgenic donor mice. *Mol. Reprod. Dev.* **53**, 142–148 (1999).
7. Nagano, M., Avarbock, M. R. & Brinster, R. L. Pattern and kinetics of mouse donor spermatogonial stem cell colonization in recipient testes. *Biol. Reprod.* **60**, 1429–1436 (1999).
8. Till, J. E. & McCulloch, E. A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* **14**, 1419–1430 (1961).
9. Shinohara, T., Orwig, K. E., Avarbock, M. R. & Brinster, R. L. Remodeling of the postnatal mouse testis is accompanied by dramatic changes in stem cell number and niche accessibility. *Proc. Natl Acad. Sci. USA* **98**, 6186–6191 (2001).
10. Ohta, H., Yomogida, K., Dohmae, K. & Nishimune, Y. Regulation of proliferation and differentiation in spermatogonial stem cells. The role of c-kit and its ligand SCF. *Development* **127**, 2025–2131 (2000).
11. Shinohara, T., Avarbock, M. R. & Brinster, R. L. β_1 - and α_6 -integrin are surface markers on mouse spermatogonial stem cells. *Proc. Natl Acad. Sci. USA* **96**, 5504–5509 (1999).
12. Lawson, K. A. *et al.* Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).

13. Ying Y, Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol. Endocrinol.* **14**, 1053–1063 (2000).
14. Wang R. A. & Zhao, G. Q. Transforming growth factor beta signal transducer Smad2 is expressed in mouse meiotic germ cells, Sertoli cells, and Leydig cells during spermatogenesis. *Biol. Reprod.* **61**, 999–1004 (1999).
15. Meng, X. *et al.* Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489–1493 (2000).
16. Tanaka, S. S. *et al.* The mouse homolog of *Drosophila vasa* is required for the development of male germ cells. *Genes Dev.* **14**, 841–853 (2000).
17. Margolis J. & Spradling, A. Identification and behavior of epithelial stem cells in the *Drosophila* ovary. *Development* **121**, 3797–3807 (1995).
18. Gonczy, P. & DiNardo, S. The germ line regulates somatic cyst cell proliferation and fate during *Drosophila* spermatogenesis. *Development* **122**, 2437–2447 (1996).
19. Xie, T. & Spradling, A. C. Decapentaplegic is essential for the maintenance and division of the germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251–260 (1998).
20. Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**, 750–754 (2000).
21. Tran, J., Brenner, T. J. & DiNardo, S. Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis. *Nature* **407**, 754–757 (2000).
22. Hardy, R. W., Tokuyasu, K. T., Lindsley, D. L. & Garavito, M. The germline proliferation center in the testis of *Drosophila melanogaster*. *J. Ultrastruct. Res.* **69**, 180–190 (1979).
23. Sahut-Barnola, I., Godt, D., Laski, F. A. & Couderc, J. L. *Drosophila* ovary morphogenesis: analysis of terminal filament formation and identification of a gene required for this process. *Dev. Biol.* **170**, 127–135 (1995).
24. Lin, H. & Spradling, A. C. Germline stem cell division and egg chamber development in transplanted *Drosophila* germlaria. *Dev. Biol.* **159**, 140–152 (1993).
25. King, F. J., Szakmary, A., Cox, D. N. & Lin, H. Yb modulates the divisions of both germline and somatic stem cells through *piwi*- and *hh*-mediated mechanisms in the *Drosophila* ovary. *Mol. Cell* **7**, 497–508 (2001).
26. de Cuevas, M. & Spradling, A. C. Morphogenesis of the *Drosophila* fusome and its implications for oocyte specification. *Development* **125**, 2781–2789 (1998).
27. Deng, W. & Lin, H. Spectrosomes and fusomes anchor mitotic spindles during asymmetric germ cell divisions and facilitate the formation of a polarized microtubule array for oocyte specification in *Drosophila*. *Dev. Biol.* **189**, 79–94 (1997).
28. Lin, H. & Spradling, A. C. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124**, 2463–2476 (1997).
29. Xie, T. & Spradling, A. C. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**, 328–330 (2000).
30. Twombly, V. *et al.* The TGF- β signalling pathway is essential for *Drosophila* oogenesis. *Development* **122**, 1555–1565 (1996).
31. Cox, D. N., Chao, A. & Lin, H. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514 (2000).
32. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
33. Moussian, B., Schoof, H., Haecker, A., Jurgens, G. & Laux, T. Role of the *ZWILLE* gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. *EMBO J.* **17**, 1799–1809 (1998).
34. Sharma, A. K. *et al.* Human CD34⁺ stem cells express the hiwi gene, a human homologue of the *Drosophila* gene *piwi*. *Blood* **97**, 426–434 (2001).
35. Forbes, A. & Lehmann, R. Nanos and Pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development* **125**, 679–690 (1998).
36. Ohlstein, B., Lavoie, C. A., Vef, O., Gateff, E. & McKearin, D. M. The *Drosophila* cystoblast differentiation factor, *benign gonial cell neoplasm*, is related to DExH-box proteins and interacts genetically with *bag-of-marbles*. *Genetics* **155**, 1809–1819 (2000).
37. Gonczy, P., Matunis, E. & DiNardo, S. *bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development* **124**, 4361–4371 (1997).
38. Henderson, S. T., Gao, D., Lambie, E. J. & Kimble, J. *lag-2* may encode a signaling ligand for the GLP-1 and LIN-12 receptors of *C. elegans*. *Development* **120**, 2913–2924 (1994).
39. Austin, J. & Kimble, J. *glp-1* is required in the germ line for regulation of the decision between mitosis and meiosis in *C. elegans*. *Cell* **51**, 589–599 (1987).
40. Berry, L. W., Westlund, B. & Schedl, T. Germ-line tumor formation caused by activation of *glp-1*, a *Caenorhabditis elegans* member of the Notch family of receptors. *Development* **124**, 925–936 (1997).
41. Kadyk, L. C., Kimble, J. Genetic regulation of entry into meiosis in *Caenorhabditis elegans*. *Development* **125**, 1803–1813 (1998).
42. Jonason, A. S. *et al.* Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc. Natl Acad. Sci. USA* **95**, 14025–14029 (1996).
43. Watt, F. in *Stem Cell Biology* (eds Marshak, D. R., Gardner, R. L. & Gottlieb, D.) 439–453 (Cold Spring Harbor Press, Cold Spring Harbor, 2001).
44. Lavker, R. M. & Sun, T.-T. Heterogeneity in epidermal basal cell keratinocytes: morphological and functional correlations. *Science* **215**, 1239–1241 (1982).
45. Cotsarelis, G., Cheng, S. Z., Dong, G., Sun, T.-T. & Lavker, R. M. Existence of slow-cycling limbal epithelial basal cells that can be preferentially stimulated to proliferate: implications on epithelial stem cells. *Cell* **57**, 201–209 (1989).
46. Rochat, A., Kobayashi, K. & Barrandon, Y. Location of stem cells of human hair follicles by clonal analysis. *Cell* **76**, 1063–1073 (1994).
47. Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T.-T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102**, 451–461 (2000).
48. Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. & Barrandon, Y. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104**, 233–245 (2001).
49. Stenn, K. S. & Paus, R. Controls of hair follicle cycling. *Physiol. Rev.* **81**, 449–494 (2001).
50. Finch, P. W., Rubin, J. S., Miki, T., Ron, D. & Aaronson, S. A. Human KGF is FGF-related with properties of a paracrine effector of epithelial cell growth. *Science* **245**, 752–755 (1989).
51. Brakebusch, C. *et al.* Skin and hair follicle integrity is crucially dependent on β 1 integrin expression on keratinocytes. *EMBO J.* **19**, 3990–4003 (2000).
52. Raghavan, S., Bauer, C., Mundscha, G., Li, Q. & Fuchs, E. Conditional ablation of beta1 integrin in skin. Severe defects in epidermal proliferation, basement membrane formation, and hair follicle invagination. *J. Cell Biol.* **150**, 1149–1160 (2000).
53. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of α -catenin in skin. *Cell* **104**, 605–617 (2001).
54. Kulssa, H., Turk, G. & Hogan, B. L. Inhibition of Bmp signalling affects growth and differentiation in the anagen hair follicle. *EMBO J.* **19**, 6664–6674 (2000).
55. Chiang, C. *et al.* Essential role for Sonic hedgehog during hair follicle morphogenesis. *Dev. Biol.* **205**, 1–9 (1999).
56. Gambardella, L., Schneider-Maunoury, S., Voiculescu, O., Charnay, P. & Barrandon, Y. Pattern of expression of the transcription factor Krox-20 in mouse hair follicle. *Mech. Dev.* **96**, 215–218 (2000).
57. van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in Lef-1-deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
58. DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
59. Huelsen, J., Vogel, R., Erdmann, B. & Cotsarelis, G. β -Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105**, 533–545 (2001).
60. Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. De novo hair follicle morphogenesis and hair tumors in mice expressing a truncated β -catenin in skin. *Cell* **95**, 605–614 (1998).
61. Chan, E. F., Gat, U., McNiff, J. M. & Fuchs, E. A common human skin tumour is caused by activating mutations in β -catenin. *Nature Genet.* **21**, 410–413 (1999).
62. Kishimoto, J., Burgeson, R. E. & Morgan, B. A. Wnt signalling maintains the hair-inducing activity of the dermal papilla. *Genes Dev.* **14**, 1181–1185 (2000).
63. Drummond-Barbosa, D. & Spradling, A. C. Stem cells and their progeny respond to nutritional changes during *Drosophila* oogenesis. *Dev. Biol.* **231**, 265–278 (2001).
64. Zhang, Y. & Kalderon, D. Hedgehog acts as a somatic stem cell factor in the *Drosophila* ovary. *Nature* **410**, 599–604 (2001).
65. Forbes, A. J., Lin, H., Ingham, P. W. & Spradling, A. C. hedgehog is required for the proliferation and specification of ovarian somatic cells prior to egg chamber formation in *Drosophila*. *Development* **122**, 1125–1135 (1996).
66. Forbes, A. J., Spradling, A. C., Ingham, P. W. & Lin, H. The role of segment polarity genes during early oogenesis in *Drosophila*. *Development* **122**, 3283–3294 (1996).
67. Zhang, Y. & Kalderon, D. Regulation of cell proliferation and patterning in *Drosophila* oogenesis by Hedgehog signaling. *Development* **127**, 2165–2176 (2000).
68. Bjerknes, M. & Cheng, H. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* **116**, 7–14 (1999).
69. Wong, M. H., Saam, J. R., Stappenbeck, T. S., Rexer, C. H. & Gordon, J. I. Genetic mosaic analysis based on Cre recombinase and navigated laser capture microdissection. *Proc. Natl Acad. Sci. USA* **97**, 12601–12606 (2000).
70. Novelli, M. R. *et al.* Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. *Science* **272**, 1187–1190 (1996).
71. Garabedian, E. M., Lisa, J. J., Roberts, M., McNevin, S. & Gordon, J. I. Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. *J. Biol. Chem.* **272**, 23729–23740 (1997).
72. Rindi, G. *et al.* Targeted ablation of secretin-producing cells in transgenic mice reveals a common differentiation pathway with multiple enteroendocrine cell lineages in the small intestine. *Development* **126**, 4149–4156 (1999).
73. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 379–383 (1998).
74. Kaestner, K. H., Silberg, D. G., Traber, P. G. & Schutz, G. The mesenchymal winged helix transcription factor Fkh6 is required for the control of gastrointestinal proliferation and differentiation. *Genes Dev.* **11**, 1583–1595 (1997).
75. Hardy, M. H. The secret life of the hair follicle. *Trends Genet.* **8**, 159–166 (1992).
76. Bjerknes, M., Cheng, H., Hay, K. & Gallinger, S. APC mutation and the crypt cycle in murine and human intestine. *Am. J. Pathol.* **150**, 833–839 (1997).
77. Wright, N. A. Epithelial stem cell repertoire in the gut: clues to the origin of cell lineages, proliferative units and cancer. *Int. J. Exp. Pathol.* **81**, 117–143 (2000).
78. van Beek, M. E. A. B., Meistrich, M. L. & de Rooij, D. G. Probability of self-renewing divisions of spermatogonial stem cells in colonies, formed after fission neutron irradiation. *Cell. Tissue Kinet.* **23**, 1–16 (1990).
79. Hebert, J. M., Rosenquist, T., Gotz, J. & Martin, G. R. FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. *Cell* **78**, 1017–1025 (1994).
80. Murillas, R. *et al.* Expression of a dominant negative mutant of epidermal growth factor receptor in the epidermis of transgenic mice elicits striking alterations in hair follicle development and skin structure. *EMBO J.* **14**, 5216–5223 (1995).
81. Sibilia, M. & Wagner, E. F. Strain-dependent epithelial defects in mice lacking the EGF receptor. *Science* **269**, 234–238 (1995).
82. Luetke, N. C. *et al.* TGF- α deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice. *Cell* **73**, 263–278 (1993).
83. Panteleyev, A. A. *et al.* Towards defining the pathogenesis of the hairless phenotype. *J. Invest. Dermatol.* **110**, 902–907 (1998).

Stem cells, cancer, and cancer stem cells

Tannishtha Reya^{*§||}, Sean J. Morrison^{†||}, Michael F. Clarke[‡] & Irving L. Weissman^{*}

^{*}Departments of Pathology and Developmental Biology, Stanford University School of Medicine, Palo Alto, California 94305, USA (e-mail: irv@stanford.edu)

[†]Howard Hughes Medical Institute, and Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109-0934, USA

[‡]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109-0936, USA

[§]Present address: Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

^{||}These authors contributed equally to this work

Stem cell biology has come of age. Unequivocal proof that stem cells exist in the haematopoietic system has given way to the prospective isolation of several tissue-specific stem and progenitor cells, the initial delineation of their properties and expressed genetic programmes, and the beginnings of their utility in regenerative medicine. Perhaps the most important and useful property of stem cells is that of self-renewal. Through this property, striking parallels can be found between stem cells and cancer cells: tumours may often originate from the transformation of normal stem cells, similar signalling pathways may regulate self-renewal in stem cells and cancer cells, and cancer cells may include 'cancer stem cells' — rare cells with indefinite potential for self-renewal that drive tumorigenesis.

Stem cells are defined as cells that have the ability to perpetuate themselves through self-renewal and to generate mature cells of a particular tissue through differentiation. In most tissues, stem cells are rare. As a result, stem cells must be identified prospectively and purified carefully in order to study their properties. Although it seems reasonable to propose that each tissue arises from a tissue-specific stem cell, the rigorous identification and isolation of these somatic stem cells has been accomplished only in a few instances. For example, haematopoietic stem cells (HSCs) have been isolated from mice and humans^{1–4}, and have been shown to be responsible for the generation and regeneration of the blood-forming and immune (haematolymphoid) systems (Fig. 1). Stem cells from a variety of organs might have the potential to be used for therapy in the future, but HSCs — the vital elements in bone-marrow transplantation — have already been used extensively in therapeutic settings (reviewed in ref. 5).

The recent discovery that bone marrow^{6–8}, as well as purified HSCs^{9,10}, can give rise to non-haematopoietic tissues suggests that these cells may have greater differentiation potential than was assumed previously. Definitive experiments are needed to determine whether the cells from the bone marrow that are capable of giving rise to different non-haematopoietic lineages are indeed HSCs or another population. If further studies support the idea of HSC plasticity, this will undoubtedly open new frontiers for understanding the developmental potential of HSCs, as well as expand their therapeutic application.

As the characteristics of HSCs, their differentiation potential and clinical applications have been covered in earlier reviews, here we discuss emerging evidence that stem cell biology could provide new insights into cancer biology. In particular, we focus on three aspects of the relationship between stem cells and tumour cells: first, the similarities in the mechanisms that regulate self-renewal of normal stem cells and cancer cells; second, the possibility that tumour cells might arise from normal stem cells; and third, the notion that

tumours might contain 'cancer stem cells' — rare cells with indefinite proliferative potential that drive the formation and growth of tumours. Through much of this review we focus on the haematopoietic system because both normal stem cells and cancer cells from this tissue are well characterized. Moreover, cancers of the haematopoietic system (that is, leukaemias) provide the best evidence that normal stem cells are the targets of transforming mutations, and that cancer cell proliferation is driven by cancer stem cells.

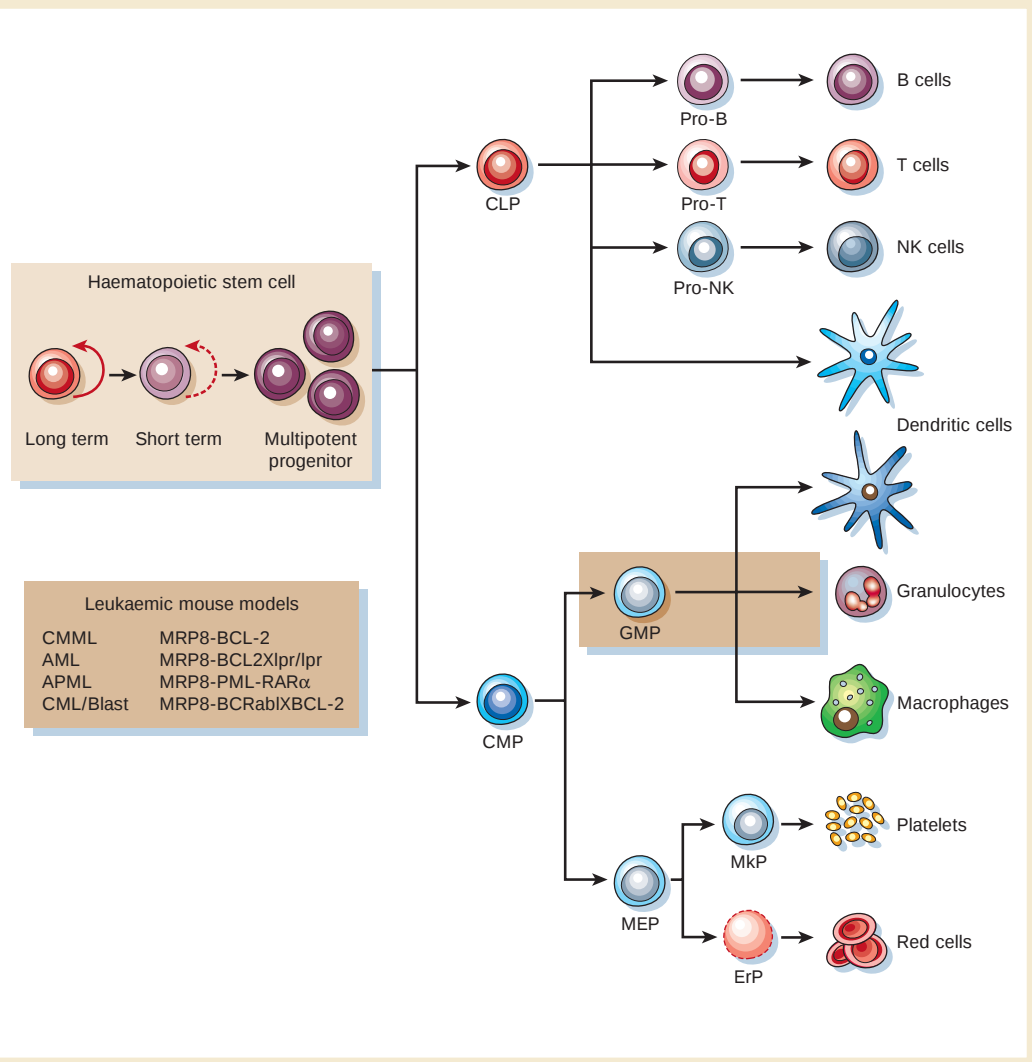
Self-renewal of haematopoietic stem cells

One of the most important issues in stem cell biology is understanding the mechanisms that regulate self-renewal. Self-renewal is crucial to stem cell function, because it is required by many types of stem cells to persist for the lifetime of the animal. Moreover, whereas stem cells from different organs may vary in their developmental potential, all stem cells must self-renew and regulate the relative balance between self-renewal and differentiation. Understanding the regulation of normal stem cell self-renewal is also fundamental to understanding the regulation of cancer cell proliferation, because cancer can be considered to be a disease of unregulated self-renewal.

In the haematopoietic system, stem cells are heterogeneous with respect to their ability to self-renew. Multipotent progenitors constitute 0.05% of mouse bone-marrow cells, and can be divided into three different populations: long-term self-renewing HSCs, short-term self-renewing HSCs, and multipotent progenitors without detectable self-renewal potential^{2,11}. These populations form a lineage in which the long-term HSCs give rise to short-term HSCs, which in turn give rise to multipotent progenitors¹¹. As HSCs mature from the long-term self-renewing pool to multipotent progenitors, they progressively lose their potential to self-renew but become more mitotically active. Whereas long-term HSCs give rise to mature haematopoietic cells for the lifetime of the mouse, short-term HSCs and multipotent progenitors reconstitute lethally irradiated mice for less than eight weeks.

Although the phenotypic and functional properties of HSCs have been extensively characterized (reviewed in

Figure 1 Development of haematopoietic stem cells. HSCs can be subdivided into long-term self-renewing HSCs, short-term self-renewing HSCs and multipotent progenitors (red arrows indicate self-renewal). They give rise to common lymphoid progenitors (CLPs; the precursors of all lymphoid cells) and common myeloid progenitors (CMPs; the precursors of all myeloid cells). Both CMPs/GMPs (granulocyte macrophage precursors) and CLPs can give rise to all known mouse dendritic cells. The isolation of precursors in the haematopoietic system has allowed the generation of a series of mouse models for myeloid leukaemia (see box, lower left). The expression of the oncogenes BCL-2, BCR-Abl and PML-RAR α under the control of the *hMRP8* promoter, individually or together, and in combination with Fas deficiency, results in diseases that resemble several human leukaemias, including chronic myelomonocytic leukaemia (CMML), acute myeloid leukaemia (AML), acute promyelocytic leukaemia (APML)⁷⁷, and chronic myeloid leukaemia (CML)/Blast (S. Jaswal, K. Akashi and I.L.W., submitted). ErP, erythrocyte precursor; MEP, megakaryocyte erythrocyte precursor; MkP, megakaryocyte precursor; NK, natural killer.



ref. 12), the fundamental question of how self-renewal is regulated remains unanswered. In most cases, combinations of growth factors that can induce potent proliferation cannot prevent the differentiation of HSCs in long-term cultures. Although progress has been made in identifying culture conditions that maintain HSC activity in culture (for example, see ref. 13), it has proved exceedingly difficult to identify combinations of defined growth factors that cause a significant expansion in culture in the number of progenitors with transplantable HSC activity.

Pathways regulating stem cell self-renewal and oncogenesis

Because normal stem cells and cancer cells share the ability to self-renew, it seems reasonable to propose that newly arising cancer cells appropriate the machinery for self-renewing cell division that is normally expressed in stem cells. Evidence shows that many pathways that are classically associated with cancer may also regulate normal stem cell development (Fig. 2). For example, the prevention of apoptosis by enforced expression of the oncogene *bcl-2* results in increased numbers of HSCs *in vivo*, suggesting that cell death has a role in regulating the homeostasis of HSCs^{14,15}.

Other signalling pathways associated with oncogenesis, such as the Notch, Sonic hedgehog (Shh) and Wnt signalling pathways, may also regulate stem cell self-renewal (reviewed in ref. 16). Notch activation in HSCs in culture using the ligand Jagged-1 have consistently increased the amount of primitive progenitor activity that can be observed *in vitro* and *in vivo*, suggesting that Notch activation promotes HSC self-renewal, or at least the maintenance of

multipotentiality^{17,18}. Shh signalling has also been implicated in the regulation of self-renewal by the finding that populations highly enriched for human HSCs (CD34⁺Lin⁻CD38⁻) exhibit increased self-renewal in response to Shh stimulation *in vitro*, albeit in combination with other growth factors¹⁹. The involvement of Notch and Shh in the self-renewal of HSCs is especially interesting in light of studies that implicate these pathways in the regulation of self-renewal of stem cells from other tissues as well (Fig. 2, and see review in this issue by Spradling and colleagues, pages 98–104).

One particularly interesting pathway that has also been shown to regulate both self-renewal and oncogenesis in different organs is the Wnt signalling pathway (Fig. 2). Wnt proteins are intercellular signalling molecules²⁰ that regulate development in several organisms²¹ and contribute to cancer when dysregulated. The expression of Wnt proteins in the bone marrow²² suggests that they may influence HSCs as well. Using highly purified mouse bone-marrow HSCs, we have shown that overexpression of activated β -catenin (a downstream activator of the Wnt signalling pathway) in long-term cultures of HSCs expands the pool of transplantable HSCs determined by both phenotype (Thy1.1^{lo}Lin^{-lo}Sca1⁺c-kit⁺) and function (ability to reconstitute the haematopoietic system *in vivo*). Moreover, ectopic expression of Axin, an inhibitor of Wnt signalling, leads to inhibition of HSC proliferation, increased death of HSCs *in vitro*, and reduced reconstitution *in vivo* (T.R. *et al.*, submitted). In separate studies, soluble Wnt proteins from conditioned supernatants have also been shown to influence the proliferation of haematopoietic progenitors from mouse fetal liver and human bone marrow^{23,24}.

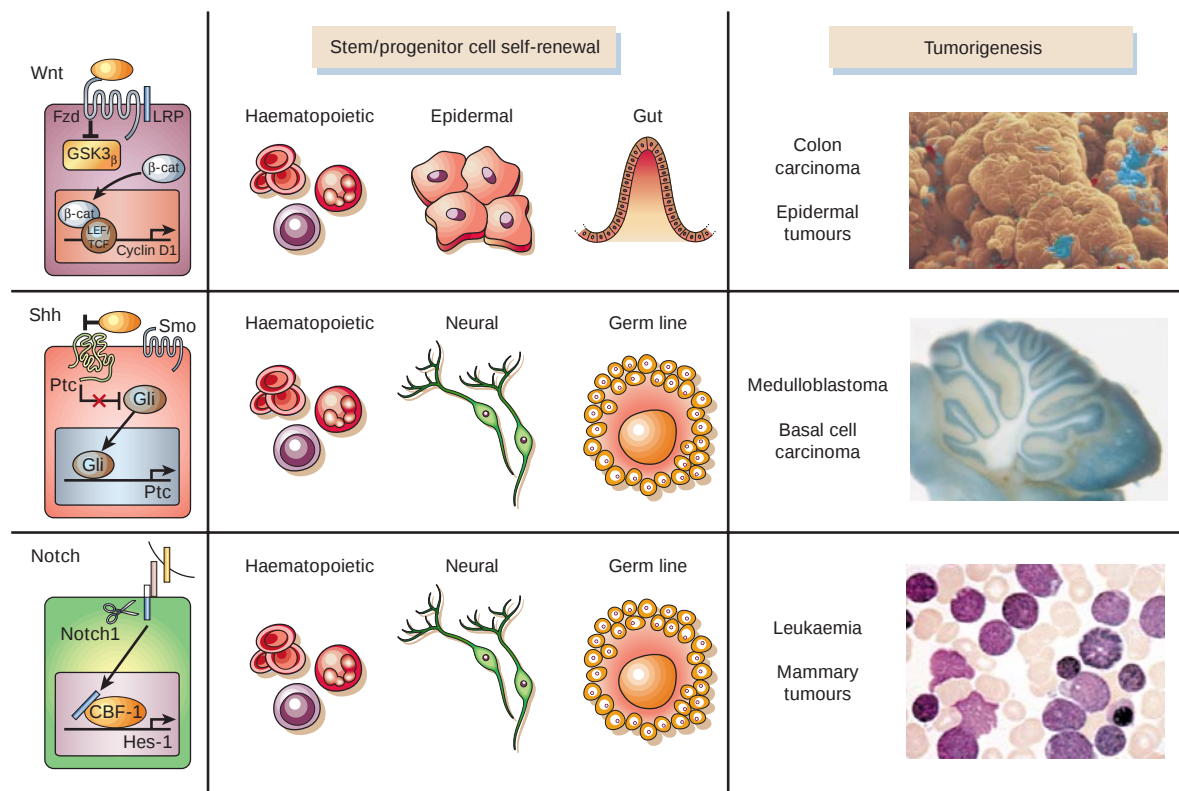


Figure 2 Signalling pathways that regulate self-renewal mechanisms during normal stem cell development and during transformation. Wnt (refs 25, 27; and T.R. *et al.*, submitted), Shh^{19,76,79} and Notch pathways^{17,80,81} have been shown to contribute to the self-renewal of stem cells and/or progenitors in a variety of organs, including the haematopoietic and nervous systems. When dysregulated, these pathways can contribute to oncogenesis. Mutations of these pathways have been associated with a number of human tumours, including colon carcinoma³⁷ and epidermal tumours⁸² (Wnt), medulloblastoma⁸³ and basal cell carcinoma⁸⁴ (Shh), and T-cell leukaemias⁸⁵ (Notch). (Images courtesy of Eye of Science/SPL and R. Wechsler-Reya/M. Scott/Annual Reviews.)

Studies of epidermal and gut progenitors suggest that the Wnt signalling pathway may contribute to the regulation of stem cell/progenitor cell self-renewal in other tissues. Cultured human keratinocytes with higher proliferative potential have increased levels of β -catenin compared with keratinocytes with lower proliferative capacity. Moreover, retroviral transduction of activated β -catenin results in increased epidermal stem cell self-renewal and decreased differentiation²⁵. *In vivo* data from transgenic mice suggest that activation of the Wnt signalling pathway in epidermal stem cells leads to epithelial cancers²⁶. Furthermore, mice lacking TCF-4, one of the transcriptional mediators of the Wnt signalling pathway, quickly exhaust the undifferentiated progenitors in the crypts of the gut epithelium during fetal development²⁷, suggesting that this pathway is required for the maintenance or self-renewal of gut epithelial stem cells.

Cumulatively, the above findings suggest that Wnt signalling may promote stem cell self-renewal in a variety of different epithelia in addition to HSCs. The molecular mechanisms by which Wnt signalling influences stem cells remain to be elucidated. It will also be important to determine whether the Wnt, Notch and Shh pathways interact to regulate stem and progenitor cell self-renewal.

Self-renewal and leukaemogenesis

If the signalling pathways that normally regulate stem cell self-renewal lead to tumorigenesis when dysregulated, then are stem cells themselves the target of transformation in certain types of cancer^{28,29}? There are two reasons to think that this may be the case. First, because stem cells have the machinery for self-renewal already activated, maintaining this activation may be simpler than turning it on *de novo*

in a more differentiated cell; that is, fewer mutations may be required to maintain self-renewal than to activate it ectopically. Second, by self-renewing, stem cells often persist for long periods of time, instead of dying after short periods of time like many mature cells in highly proliferative tissues. This means that there is a much greater opportunity for mutations to accumulate in individual stem cells than in most mature cell types (Fig. 3).

Even restricted progenitor cells are less likely than stem cells to undergo neoplastic transformation because they proliferate for a much shorter period of time before terminally differentiating. Restricted haematopoietic progenitors of the lymphoid³⁰ and myeloid lineages all fail to self-renew detectably on transplantation (K. Nankorn, Traver, D., I.L.W. and K. Akashi, submitted). Thus, restricted progenitors would first need to acquire the extensive self-renewal potential of stem cells to have the opportunity to experience additional mutations that would lead to transformation. Nonetheless, restricted progenitors could potentially be transformed either by acquiring mutations that cause them to self-renew like stem cells, or by inheriting existing mutations from stem cells such that only a single mutation is required in the progenitors to cause transformation (Fig. 3).

Stem cells as targets of mutation

For most cancers, the target cell of transforming mutations is unknown; however, there is considerable evidence that certain types of leukaemia arise from mutations that accumulate in HSCs. The cells capable of initiating human acute myeloid leukaemia (AML) in NOD/SCID (non-obese diabetic/severe combined immunodeficiency) mice have a CD34⁺CD38⁺ phenotype in most AML subtypes, and thus have a phenotype similar to normal HSCs³¹. Conversely,

CD34⁺CD38⁺ leukaemia cells cannot transfer disease to mice in the vast majority of cases, despite the fact that they exhibit a leukaemic blast phenotype. This suggests that normal HSCs rather than committed progenitors are the target for leukaemic transformation.

The most frequent chromosomal abnormalities in AML involve the 8;21 translocation, which results in AML1-ETO chimaeric transcripts in leukaemic cells. In work done on human HSCs from patients in remission, AML1-ETO transcripts were found in a fraction of normal HSCs in the marrow³². These prospectively isolated HSCs and their progeny were not leukaemic, and could differentiate to normal myeloerythroid cells *in vitro*. This indicates that the translocation occurred originally in normal HSCs and that additional mutations in a subset of these HSCs or their progeny subsequently lead to leukaemia³². In this study, the normal HSCs were CD34⁺CD38⁺Thy-1⁺, whereas the leukaemic blasts were CD34⁺CD38⁺Thy-1⁻. Although the translocation must have occurred in normal HSCs, subsequent transforming mutations might have occurred either in downstream Thy-1⁻ progenitors, or in HSCs if one consequence of neoplastic proliferation was the loss of Thy-1 expression. The idea that stem cells are a common target of pre-leukaemic events or leukaemic transformation is also supported by work in lymphoid³³ and chronic myeloid leukaemias³⁴ where clonotypic leukaemia-associated chromosomal rearrangements have also been found in CD34⁺CD38⁺ cells, a population enriched for HSCs. Thus, a variety of leukaemias may arise from mutations that accumulate in HSCs to cause their malignant transformation at the stage of stem cells or their progeny.

Progenitor cells as targets of transformation

Although stem cells are often the target of genetic events that are necessary or sufficient for malignant transformation, in other cases restricted progenitors or even differentiated cells may become transformed (Fig. 3). By targeting the expression of transgenes specifically to restricted myeloid progenitors using the *hMRP-8* promoter, it is

possible to create a mouse model in which myeloid leukaemia arises from restricted progenitors. These leukaemias resemble human leukaemias in many respects, even though the targeted genetic changes cause the leukaemias to arise from restricted progenitors rather than stem cells. For example, we have generated transgenic mouse models for myeloid leukaemias using an *hMRP-8* promoter, which targets the expression of transgenes specifically to myeloid progenitors³⁵. The enforced expression of the anti-apoptotic gene *bcl-2* in the myeloid lineage leads to a disease that is similar to human chronic myelomonocytic leukaemia, including monocytosis, splenomegaly and neutropenia, as the mice age. However, these mice rarely develop acute malignancies.

To test whether additional mutations are required to synergize with *bcl-2* to promote AML, *hMRP8-bcl-2* transgenic mice were bred with *lpr/lpr* Fas-deficient mice. Remarkably, the loss of these two distinct apoptosis pathways led to the development of AML in 15% of the mice³⁶. These mice have an expansion of myeloblasts in all haematopoietic tissues, with a substantially lowered number of granulocytes in the marrow and blood. These studies show that prevention of cell death is a crucial event in myeloid leukaemogenesis and that restricted progenitors can be transformed. As described above, in the case of spontaneously arising human leukaemias it is likely that stem cells accumulate the mutations that are necessary for neoplastic proliferation; however, these mutations may accumulate in stem cells even while the effects of the mutations are expressed in restricted progenitors. That is, mutations that accumulate in stem cells may lead to neoplastic proliferation of primitive progenitors downstream of stem cells.

Perhaps the reason why only 15% of mice progress to AML in mice expressing *Bcl-2* and lacking *Fas* is that the progenitors in these mice also must acquire an additional mutation that causes dysregulated self-renewal (Fig. 3). If a single additional mutation causes transformation then this transforming event is probably a gain-of-function mutation, such as one that promotes constitutive self-renewal. Because stabilized β -catenin can promote the self-renewal of HSCs and other types of progenitors (ref. 25, and T.R. *et al.*, submitted; Fig. 2), we propose that gain-of-function mutations in β -catenin may, in many cases, transform deathless pre-malignant cells to cancer cells by promoting proliferation. In support of this is evidence to show that activation of β -catenin and dysregulation of the Wnt signalling pathway in general is common in cancer³⁷, and that the targeted overactivation of this pathway can lead to tumours in transgenic mice³⁸. It is also possible that mutations in other signalling pathways promote progenitor self-renewal. It is important to study this further, because understanding the molecular basis of the unregulated self-renewal of cancer cells will allow the design of more effective therapies.

In essence, newly arising cancer cells may appropriate the machinery for self-renewing cell divisions that is normally expressed in stem cells. In the haematopoietic system, the only long-term self-renewing cells in the myeloerythroid pathway (Fig. 1, bottom) are HSCs; however, at least two differentiated cell types (Fig. 1, top) can also self-renew. Both T and B lymphocytes undergo clonal expansion on stimulation to produce resting memory lymphocytes. These lymphocytes proliferate again when the antigens are re-encountered. Lymphoid leukaemias can activate these receptor-mediated mitogenic pathways in the course of leukaemogenesis³⁹⁻⁴³.

Cancer stem cells and aberrant organogenesis

Basic cancer research has focused on identifying the genetic changes that lead to cancer. This has led to major advances in our understanding of the molecular and biochemical pathways that are involved in tumorigenesis and malignant transformation. But while we have focused on the molecular biology of cancer, our understanding of the cellular biology has lagged. That is, although we understand (to a first approximation) the effects of particular mutations on the proliferation and survival of model cells, such as fibroblasts or cell lines, we

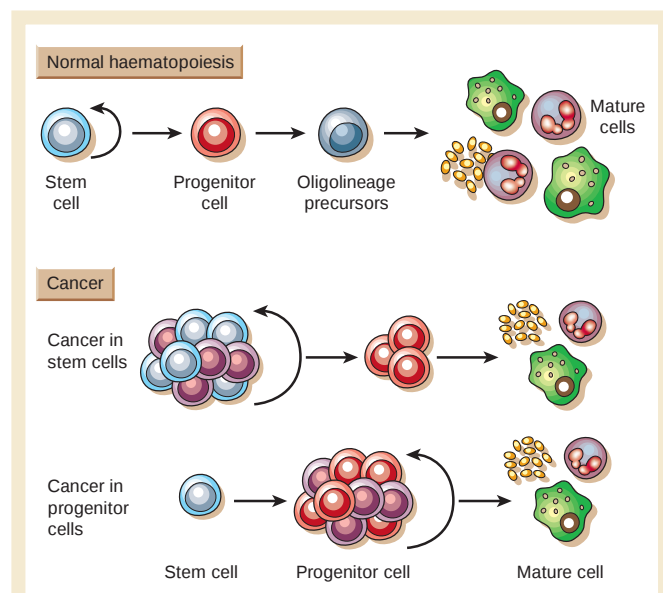


Figure 3 Comparison of self-renewal during haematopoietic stem cell development and leukaemic transformation. Because of their high level of self-renewal, stem cells are particularly good targets of leukaemic transformation. Unlike normal haematopoiesis, where signalling pathways that have been proposed to regulate self-renewal are tightly regulated (top), during transformation of stem cells, the same mechanisms may be dysregulated to allow uncontrolled self-renewal (middle). Furthermore, if the transformation event occurs in progenitor cells, it must endow the progenitor cell with the self-renewal properties of a stem cell, because these progenitors would otherwise differentiate (bottom).

can often only guess what the effects of such mutations will be on the actual cells involved in particular cancers. This has handicapped our ability to translate our identification of mutations into new therapies.

A tumour can be viewed as an aberrant organ initiated by a tumorigenic cancer cell that acquired the capacity for indefinite proliferation through accumulated mutations. If one views a tumour as an abnormal organ, then the principles of normal stem cell biology^{12,44} can be applied to understand better how tumours develop (reviewed in ref. 45). In fact, many observations suggest that analogies between normal stem cells and tumorigenic cells may be appropriate. Both normal stem cells and tumorigenic cells have extensive proliferative potential and the ability to give rise to new (normal or abnormal) tissues. Both tumours and normal tissues are composed of heterogeneous combinations of cells, with different phenotypic characteristics and different proliferative potentials^{46–49}. Because most tumours have a clonal origin^{50–52}, tumorigenic cancer cells must give rise to phenotypically diverse progeny, including cancer cells with indefinite proliferative potential, as well as cancer cells with limited or no proliferative potential. This suggests that tumorigenic cancer cells undergo processes that are analogous to the self-renewal and differentiation of normal stem cells.

Although some of the heterogeneity in tumours arises as a result of continuing mutagenesis, it is likely that heterogeneity also arises through the aberrant differentiation of cancer cells. It is well documented that many types of tumours contain cancer cells with heterogeneous phenotypes reflecting aspects of the differentiation that normally occurs in the tissues from which the tumours arise. The variable expression of normal differentiation markers by cancer cells in a tumour suggests that some of the heterogeneity in tumours arises as a result of the anomalous differentiation of tumour cells. Examples of this include the variable expression of myeloid markers in chronic myeloid leukaemia, the variable expression of neuronal markers within peripheral neuroectodermal tumours, and the variable expression of milk proteins or the oestrogen receptor within breast cancer.

In other words, both normal stem cells and tumorigenic cells give rise to phenotypically heterogeneous cells that exhibit various degrees of differentiation. Thus, tumorigenic cells can be thought of as cancer stem cells that undergo an aberrant and poorly regulated process of organogenesis analogous to what normal stem cells do. It is perhaps not surprising that tumorigenic cells behave in ways that are analogous to normal stem cells given that cancer cells tend to display functional and phenotypic attributes of the normal cells from which they are derived²⁸.

Evidence for cancer stem cells

It was first extensively documented for leukaemia and multiple myeloma that only a small subset of cancer cells is capable of extensive proliferation. For example, when mouse myeloma cells were obtained from mouse ascites, separated from normal haematopoietic cells and put in clonal *in vitro* colony-forming assays, only 1 in 10,000 to 1 in 100 cancer cells were able to form colonies⁵³. Even when leukaemic cells were transplanted *in vivo*, only 1–4% of cells could form spleen colonies^{54–56}. Because the differences in clonogenicity among the leukaemia cells mirrored the differences in clonogenicity among normal haematopoietic cells, the clonogenic leukaemic cells were described as leukaemic stem cells (for example, see ref. 53). But two formal possibilities remained: either all leukaemia cells had a low probability of proliferating extensively in these assays such that all leukaemia cells had the potential to behave as leukaemic stem cells, or most leukaemia cells were unable to proliferate extensively and only a small, definable subset of cells was consistently clonogenic.

To prove the second possibility, it would be necessary to separate different classes of leukaemia cell and show that one subset is highly enriched for clonogenic capacity and all other cells are greatly depleted for clonogenicity. This has been accomplished by Dick and

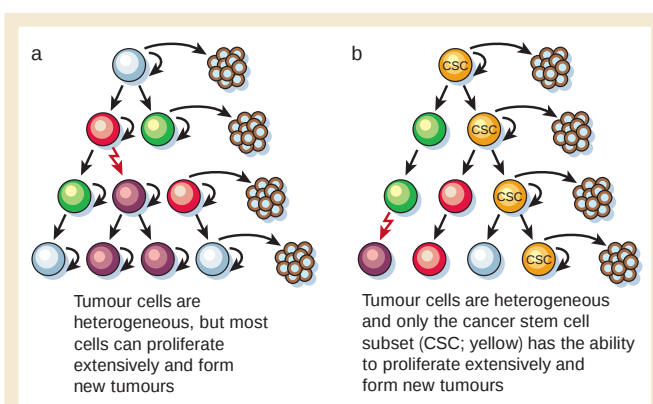


Figure 4 Two general models of heterogeneity in solid cancer cells. **a**, Cancer cells of many different phenotypes have the potential to proliferate extensively, but any one cell would have a low probability of exhibiting this potential in an assay of clonogenicity or tumorigenicity. **b**, Most cancer cells have only limited proliferative potential, but a subset of cancer cells consistently proliferate extensively in clonogenic assays and can form new tumours on transplantation. The model shown in **b** predicts that a distinct subset of cells is enriched for the ability to form new tumours, whereas most cells are depleted of this ability. Existing therapeutic approaches have been based largely on the model shown in **a**, but the failure of these therapies to cure most solid cancers suggests that the model shown in **b** may be more accurate.

colleagues⁵⁷, who showed that human AML stem cells could be identified prospectively and purified as CD34⁺CD38[−] cells from patient samples. Despite the fact that these cells represented a small but variable proportion of AML cells (0.2% in one patient), they were the only cells capable of transferring AML from human patients to NOD/SCID mice in the vast majority of cases. This excluded the first possibility that all AML cells had a similar clonogenic capacity, and showed that a small, predictable subset was consistently enriched for the ability to proliferate and transfer disease.

It has also been shown for solid cancers that the cells are phenotypically heterogeneous and that only a small proportion of cells are clonogenic in culture and *in vivo*^{46–49,58}. For example, only 1 in 1,000 to 1 in 5,000 lung cancer, ovarian cancer or neuroblastoma cells were found to form colonies in soft agar⁵⁹. Just as in the context of leukaemic stem cells, these observations led to the hypothesis that only a few cancer cells are actually tumorigenic and that these tumorigenic cells could be considered as cancer stem cells⁵⁹. But, as explained above, two possibilities remain: either all solid cancer cells have a low probability of proliferating extensively and behaving in clonogenic assays as cancer stem cells, or most cancer cells have only a limited proliferative potential and cannot behave as cancer stem cells, but a small, definable subset of cells is enriched for the ability to proliferate extensively and form tumours.

In both cases, some of the cancer cell heterogeneity would arise as a result of environmental differences within the tumour and continuing mutagenesis. The essential difference between these possibilities is the prediction, according to the second possibility, that whatever the environment or mutational status of the cells, only a small, phenotypically distinct subset of cancer cells has the ability to proliferate extensively or form a new tumour (Fig. 4). It has not been possible to distinguish between these models of solid cancer heterogeneity, because as yet no one has published the identity of purified subsets of uncultured solid cancer cells that are enriched for the ability to form new tumours.

The implications of solid cancer stem cells

If the growth of solid cancers were driven by cancer stem cells, it would have profound implications for cancer therapy. At present, all of the phenotypically diverse cancer cells are treated as though they

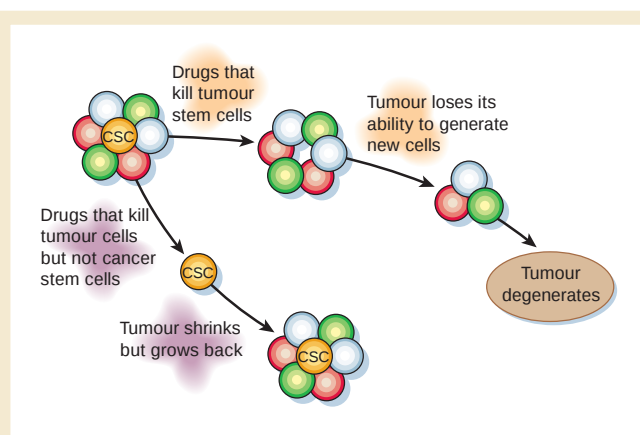


Figure 5 Conventional therapies may shrink tumours by killing mainly cells with limited proliferative potential. If the putative cancer stem cells are less sensitive to these therapies, then they will remain viable after therapy and re-establish the tumour. By contrast, if therapies can be targeted against cancer stem cells, then they might more effectively kill the cancer stem cells, rendering the tumours unable to maintain themselves or grow. Thus, even if cancer stem cell-directed therapies do not shrink tumours initially, they may eventually lead to cures.

have unlimited proliferative potential and can acquire the ability to metastasize. For many years, however, it has been recognized that small numbers of disseminated cancer cells can be detected at sites distant from primary tumours in patients that never manifest metastatic disease^{58,60}. One possibility is that immune surveillance is highly effective at killing disseminated cancer cells before they can form a detectable tumour. Another possibility is that most cancer cells lack the ability to form a new tumour such that only the dissemination of rare cancer stem cells can lead to metastatic disease (reviewed in ref. 45). If so, the goal of therapy must be to identify and kill this cancer stem cell population. If solid cancer stem cells can be identified prospectively and isolated, then we should be able to identify more efficiently new diagnostic markers and therapeutic targets expressed by the stem cells.

If tumour growth and metastasis are driven by a small population of cancer stem cells, this might explain the failure to develop therapies that are consistently able to eradicate solid tumours⁶¹. Although currently available drugs can shrink metastatic tumours, these effects are usually transient and often do not appreciably extend the life of patients^{62,63}. One reason for the failure of these treatments is the acquisition of drug resistance by the cancer cells as they evolve; another possibility is that existing therapies fail to kill cancer stem cells effectively.

Existing therapies have been developed largely against the bulk population of tumour cells because they are often identified by their ability to shrink tumours. Because most cells with a cancer have limited proliferative potential, an ability to shrink a tumour mainly reflects an ability to kill these cells. It seems that normal stem cells from various tissues tend to be more resistant to chemotherapeutics than mature cell types from the same tissues⁶⁴. The reasons for this are not clear, but may relate to high levels of expression of anti-apoptotic proteins^{65–68} or ABC transporters such as the multidrug resistance gene^{69,70}. If the same were true of cancer stem cells, then one would predict that these cells would be more resistant to chemotherapeutics than tumour cells with limited proliferative potential. Even therapies that cause complete regression of tumours might spare enough cancer stem cells to allow regrowth of the tumours. Therapies that are more specifically directed against cancer stem cells might result in much more durable responses and even cures of metastatic tumours (Fig. 5).

Genomics may provide a powerful means for identifying drug targets in cancer cells. Although targeting genetic mutations does not

require isolation of the stem cells, there are likely to be differences in gene expression between cancer stem cells and tumour cells with limited proliferative potential. The application of microarray analysis to malignant tumours has shown that patterns of gene expression can be used to group tumours into different categories, often reflecting different mutations^{71–74}. As a result, tumour types that cannot be distinguished pathologically, but that can be distinguished on the basis of differences in gene-expression profile, can be examined for differences in treatment sensitivity. However, gene-expression profiling is often conducted on tumour samples that contain a mixture of normal cells, highly proliferative cancer cells, and cancer cells with limited proliferation potential. This results in a composite profile that may obscure differences between tumours, because the highly proliferative cells that drive tumorigenesis often represent a minority of cancer cells. Gene-expression profiling of cancer stem cells would allow the profile to reflect the biology of the cells that are actually driving tumorigenesis. Microdissection of morphologically homogeneous collections of cancer cells is one way of generating profiles that reflect more homogeneous collections of cells^{75,76}. The next frontier will be to purify the cancer stem cells from the whole tumour that retain unlimited proliferative potential and to perform gene-expression profiling on those cells. In addition to being a more efficient way of identifying new therapeutic and diagnostic targets, the profiling of cancer stem cells might sharpen the differences in patterns observed between different tumours.

Perspectives

The ideas discussed in this review can be summarized as a set of propositions. First, self-renewal is the hallmark property of stem cells in normal and neoplastic tissues. Second, in the haematopoietic system, long-term self-renewal is limited to rare long-term HSCs and some lymphocytes; other cell types lack this potential. Third, cells that continue to divide over long periods of time are much more likely to accumulate mutations that cause neoplasia. Thus genetic changes that lead to myeloid leukaemias must occur either in long-term HSCs or in progeny that first acquire the ability to self-renew. The fact that normal long-term HSCs in leukaemia patients often have leukaemia-associated translocations strongly supports the idea that leukaemic mutations often accumulate in HSCs. Mutations that lead to certain types of lymphoma may accumulate in lymphocytes, given their ability to self-renew over the long term. Fourth, in other normal tissues that contain self-renewing stem cells, such as the epithelia, the genetic changes that are steps in the progression to solid tumours probably also occur in the stem cells, or in progeny that acquire the potential for self-renewal. Fifth, distinct signalling pathways control stem cell self-renewal in different tissues. But perhaps within individual tissues, the same pathways are used consistently by both normal stem cells and cancer cells to regulate proliferation. For example, Wnt signalling regulates the self-renewal of normal stem cells in the blood and epithelia. Constitutive activation of the Wnt pathway has been implicated in a number of epithelial cancers. The regulation and consequences of Wnt signalling in normal and neoplastic cells need to be further elucidated. Sixth, understanding the signalling pathways that are used by for normal stem cells and neoplastic cells should facilitate the use of normal stem cells for regenerative medicine and the identification of cancer stem cell targets for anticancer therapies. Seventh, within most tumours there may exist cancer stem cells that can self-renew indefinitely, in contrast to most stem cells that may have limited proliferative potential. Finally, in order to cure cancer, it is necessary and sufficient to kill cancer stem cells. To accomplish this it will be necessary to identify and characterize the properties of these cells.

There are many connections between stem cells and cancer that are important to understand. Just as the signals that are known to control oncogenesis are providing clues about the control of

self-renewal of normal stem cells, studies of stem cell biology are lending insight into the origins of cancer and will ultimately yield new approaches to fight this disease. □

1. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
2. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
3. Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M. & Peault, B. Isolation of a candidate human hematopoietic stem-cell population. *Proc. Natl Acad. Sci. USA* **89**, 2804–2808 (1992).
4. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
5. Akashi, K. & Weissman, I. L. In *Developmental Biology of Hematopoiesis* (ed. Zon, L. I.) 15–34 (Oxford Univ. Press, New York, 2001).
6. Petersen, B. E. *et al.* Bone marrow as a potential source of hepatic oval cells. *Science* **284**, 1168–1170 (1999).
7. Brazelton, T. R., Rossi, F. M. V., Keshet, G. I. & Blau, H. M. From marrow to brain: expression of neuronal phenotypes in adult mice. *Science* **290**, 1775–1779 (2000).
8. Mezey, E., Chandross, K. J., Harta, G., Maki, R. A. & McKeercher, S. R. Turning blood into brain: cells bearing neuronal antigens generated *in vivo* from bone marrow. *Science* **290**, 1779–1782 (2000).
9. Lagasse, E. *et al.* Purified hematopoietic stem cells can differentiate to hepatocytes *in vivo*. *Nature Med.* **6**, 1229–1234 (2000).
10. Krause, D. S. *et al.* Multi-organ, multi-lineage engraftment by a single bone marrow derived stem cell. *Cell* **105**, 369–377 (2001).
11. Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E. & Weissman, I. L. Identification of a lineage of multipotent hematopoietic progenitors. *Development* **124**, 1929–1939 (1997).
12. Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
13. Miller, C. L. & Eaves, C. J. Expansion *in vitro* of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability. *Proc. Natl Acad. Sci. USA* **94**, 13648–13653 (1997).
14. Domen, J., Gandy, K. L. & Weissman, I. L. Systemic overexpression of BCL-2 in the hematopoietic system protects transgenic mice from the consequences of lethal irradiation. *Blood* **91**, 2272–2282 (1998).
15. Domen, J. & Weissman, I. L. Hematopoietic stem cells need two signals to prevent apoptosis; BCL-2 can provide one of these, Kit/c-Kit signaling the other. *J. Exp. Med.* **192**, 1707–1718 (2000).
16. Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signaling pathways in cancer. *Nature* **411**, 349–354 (2001).
17. Varnum-Finney, B. *et al.* Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. *Nature Med.* **6**, 1278–1281 (2000).
18. Karanu, F. N. *et al.* The Notch ligand Jagged-1 represents a novel growth factor of human hematopoietic stem cells. *J. Exp. Med.* **192**, 1365–1372 (2000).
19. Bhardwaj, G. *et al.* Sonic hedgehog induces the proliferation of primitive human hematopoietic cells via BMP regulation. *Nature Immunol.* **2**, 172–180 (2001).
20. Nusse, R. & Varmus, H. E. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* **31**, 99–109 (1982).
21. Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
22. Reya, T. *et al.* Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity* **13**, 15–24 (2000).
23. Austin, T. W., Solar, G. P., Ziegler, F. C., Liem, L. & Matthews, W. A role for the Wnt gene family in hematopoiesis: expansion of multilineage progenitor cells. *Blood* **89**, 3624–3635 (1997).
24. Van Den Berg, D. J., Sharma, A. K., Bruno, E. & Hoffman, R. Role of members of the Wnt gene family in human hematopoiesis. *Blood* **92**, 3189–3202 (1998).
25. Zhu, A. J. & Watt, F. M. β -catenin signalling modulates proliferative potential of human epidermal keratinocytes independently of intercellular adhesion. *Development* **126**, 2285–2298 (1999).
26. Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. De novo hair follicle morphogenesis and hair tumors in mice expressing a truncated β -catenin in skin. *Cell* **95**, 605–614 (1998).
27. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 1–5 (1998).
28. Sell, S. & Pierce, G. B. Maturation arrest of stem cell differentiation is a common pathway for the cellular origin of teratocarcinomas and epithelial cancers. *Lab. Invest.* **70**, 6–22 (1994).
29. Sawyers, C., Denny, C. & Witte, O. Leukemia and the disruption of normal hematopoiesis. *Cell* **64**, 337–350 (1991).
30. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
31. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
32. Miyamoto, T., Weissman, I. L. & Akashi, K. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc. Natl Acad. Sci. USA* **97**, 7521–7526 (2000).
33. George, A. A. *et al.* Detection of leukemic cells in the CD34⁺CD38⁺ bone marrow progenitor population in children with acute lymphoblastic leukemia. *Blood* **97**, 3925–3930 (2001).
34. Mauro, M. J. & Druker, B. J. Chronic myelogenous leukemia. *Curr. Opin. Oncol.* **13**, 7–17 (2001).
35. Lagasse, E. & Weissman, I. L. bcl-2 inhibits apoptosis of neutrophils but not their engulfment by macrophages. *J. Exp. Med.* **179**, 1047–1052 (1994).
36. Traver, D., Akashi, K., Weissman, I. L. & Lagasse, E. Mice defective in two apoptosis pathways in the myeloid lineage develop acute myeloblastic leukemia. *Immunity* **9**, 47–57 (1998).
37. Polakis, P. Wnt signaling and cancer. *Genes Dev.* **14**, 1837–1851 (2000).
38. Tsukamoto, A., Grosschedl, R., Guzman, R., Parslow, T. & Varmus, H. E. Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell* **55**, 619–625 (1988).
39. Weissman, I. L. & Baird, S. In *Neoplastic Transformation: Mechanisms and Consequences* (ed. Korprowski, H.) 135–152 (Dahlem Konferenzen, Berlin, 1977).
40. McGrath, M. S. & Weissman, I. L. AKR leukemogenesis: identification and biological significance of thymic lymphoma receptors for AKR retroviruses. *Cell* **17**, 65–75 (1979).
41. McGrath, M. S., Pillemer, E. & Weissman, I. L. Murine leukaemogenesis: monoclonal antibodies to T-cell determinants arrest T-lymphoma cell proliferation. *Nature* **285**, 259–261 (1980).
42. Quinn, E. R. *et al.* The B cell receptor of a hepatitis C virus associated non-Hodgkin's lymphoma binds the viral E-2 envelope protein, implicating immunoglobulin activation in lymphomagenesis. *Blood* (in press).
43. O'Neill, H. C., McGrath, M. S., Allison, J. P. & Weissman, I. L. A subset of T cell receptors associated with L3T4 molecules mediates CGVL leukemia cell binding of its cognate retrovirus. *Cell* **49**, 143–151 (1987).
44. Morrison, S. J., Shah, N. M. & Anderson, D. J. Regulatory mechanisms in stem cell biology. *Cell* **88**, 287–298 (1997).
45. Kummermehr, J. & Trott, K.-R. In *Stem Cells* (ed. Potten, C. S.) 363–399 (Academic, New York, 1997).
46. Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
47. Fidler, I. J. & Hart, I. R. Biological diversity in metastatic neoplasms: origins and implications. *Science* **217**, 998–1003 (1982).
48. Heppner, G. H. Tumor heterogeneity. *Cancer Res.* **44**, 2259–2265 (1984).
49. Nowell, P. C. Mechanisms of tumor progression. *Cancer Res.* **46**, 2203–2207 (1986).
50. Nowell, P. C. A minute chromosome in human granulocytic leukemia. *Science* **132**, 1497 (1960).
51. Fialkow, P. J. Clonal origin of human tumors. *Biochim. Biophys. Acta* **458**, 283–321 (1976).
52. Fearon, E. R., Hamilton, S. R. & Vogelstein, B. Clonal analysis of human colorectal tumors. *Science* **238**, 193–197 (1987).
53. Park, C. H., Bergsagel, D. E. & McCulloch, E. A. Mouse myeloma tumor stem cells: a primary cell culture assay. *J. Natl Cancer Inst.* **46**, 411–422 (1971).
54. Bruce, W. R. & Gaag, H. v. d. A quantitative assay for the number of murine lymphoma cells capable of proliferation *in vivo*. *Nature* **199**, 79–80 (1963).
55. Wodinsky, I., Swinarski, J. & Kensler, C. J. Spleen colony studies of leukemia L1210. I. Growth kinetics of lymphocytic L1210 cells *in vivo* as determined by spleen colony assay. *Cancer Chemother. Rep.* **51**, 415–421 (1967).
56. Bergsagel, D. E. & Valeriote, F. A. Growth characteristics of a mouse plasma cell tumor. *Cancer Res.* **28**, 2187–2196 (1968).
57. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
58. Southam, C. M. & Brunschwig, A. Quantitative studies of autotransplantation of human cancer. *Cancer* **14**, 971–978 (1961).
59. Hamburger, A. W. & Salmon, S. E. Primary bioassay of human tumor stem cells. *Science* **197**, 461–463 (1977).
60. Salsbury, A. J. The significance of the circulating cancer cell. *Cancer Treatment Rev.* **2**, 55–72 (1975).
61. Williams, S. D. Treatment of disseminated germ cell tumors with cisplatin, bleomycin, and either vinblastine or etoposide. *N. Engl. J. Med.* **316**, 1435–1439 (1987).
62. Stockler, M., Wilcken, N. R. C., Ghersi, D. & Simes, R. J. Systematic reviews of chemotherapy and endocrine therapy in metastatic breast cancer. *Cancer Treatment Rev.* **26**, 151–168 (2000).
63. Lippman, M. E. High-dose chemotherapy plus autologous bone marrow transplantation for metastatic breast cancer. *N. Engl. J. Med.* **342**, 1119–1120 (2000).
64. Harrison, D. E. & Lerner, C. P. Most primitive hematopoietic stem cells are stimulated to cycle rapidly after treatment with 5-fluorouracil. *Blood* **78**, 1237–1240 (1991).
65. Bouwens, L. & DeBlay, E. Islet morphogenesis and stem cell markers in rat pancreas. *J. Histochem. Cytochem.* **44**, 947–951 (1996).
66. Peters, R., Leyvraz, S. & Perey, L. Apoptotic regulation in primitive hematopoietic precursors. *Blood* **92**, 2041–2052 (1998).
67. Domen, J., Gandy, K. L. & Weissman, I. L. Systemic overexpression of BCL-2 in the hematopoietic system protects transgenic mice from the consequences of lethal irradiation. *Blood* **91**, 2272–2282 (1998).
68. Feuerhake, F., Sigg, W., Hofer, E. A., Dimpfl, T. & Welsch, U. Immunohistochemical analysis of Bcl-2 and Bax expression in relation to cell turnover and epithelial differentiation markers in the non-lactating human mammary gland epithelium. *Cell Tissue Res.* **299**, 47–58 (2000).
69. Zhou, S. *et al.* The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nature Med.* **7**, 1028–1034 (2001).
70. Terskikh, A. V. *et al.* From hematopoiesis to neurogenesis: evidence of overlapping genetic programs. *Proc. Natl Acad. Sci. USA* **98**, 7934–7939 (2001).
71. Bittner, M. *et al.* Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**, 536–540 (2000).
72. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
73. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
74. Golub, T. R. Genome-wide views of cancer. *N. Engl. J. Med.* **344**, 601–602 (2001).
75. Sgroi, D. C. *et al.* *In vivo* gene expression profile analysis of human breast cancer progression. *Cancer Res.* **59**, 5656–5661 (1999).
76. Leethanakul, C. *et al.* Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene* **19**, 3220–3224 (2000).
77. Kogan, S. C. *et al.* BCL-2 cooperates with promyelocytic leukemia retinoic acid receptor α chimeric protein (PMLRAR α) to block neutrophil differentiation and initiate acute leukemia. *J. Exp. Med.* **193**, 531–543 (2001).
78. Wechsler-Reya, R. J. & Scott, M. P. Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron* **22**, 103–114 (1999).
79. Zhang, Y. & Kalderon, D. Hedgehog acts as a somatic stem cell factor in the *Drosophila* ovary. *Nature* **410**, 599–604 (2001).
80. Henrique, D. *et al.* Maintenance of neuroepithelial progenitor cells by Delta-Notch signalling in the embryonic chick retina. *Curr. Biol.* **7**, 661–670 (1997).
81. Austin, J. & Kimble, J. glp-1 is required in the germ line for regulation of the decision between mitosis and meiosis in *C. elegans*. *Cell* **51**, 589–599 (1987).
82. Chan, E. F., Gat, U., McNiff, J. M. & Fuchs, E. A common human skin tumour is caused by activating mutations in β -catenin. *Nature Genet.* **21**, 410–413 (1999).
83. Wechsler-Reya, R. & Scott, M. P. The developmental biology of brain tumors. *Annu. Rev. Neurosci.* **24**, 385–428 (2001).
84. Gailani, M. R. & Bale, A. E. Acquired and inherited basal cell carcinomas and the patched gene. *Adv. Dermatol.* **14**, 261–283 (1999).
85. Ellisen, L. W. *et al.* TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* **66**, 649–661 (1991).

The development of neural stem cells

Sally Temple

Center for Neuropharmacology and Neuroscience, Albany Medical College, 47 New Scotland Avenue, Albany, New York 12208, USA
(e-mail: temples@mail.amc.edu)

The discovery of stem cells that can generate neural tissue has raised new possibilities for repairing the nervous system. A rush of papers proclaiming adult stem cell plasticity has fostered the notion that there is essentially one stem cell type that, with the right impetus, can create whatever progeny our heart, liver or other vital organ desires. But studies aimed at understanding the role of stem cells during development have led to a different view — that stem cells are restricted regionally and temporally, and thus not all stem cells are equivalent. Can these views be reconciled?

The discovery of neural stem cells was rooted in classic studies of haematopoiesis and of invertebrate neural development, which inspired examination of single neural progenitor cells. (In this review, I will use the term 'progenitor cell' to refer to all classes of immature, proliferating cells. Neural stem cells are a subtype of progenitor cells in the nervous system that can self-renew and generate both neurons and glia.) The early studies led to the isolation of stem-like cells from the embryonic mammalian central nervous system (CNS)¹⁻⁴ and the peripheral nervous system (PNS)⁵. Since then, stem cells have been isolated from many regions of the embryonic nervous system, indicating their ubiquity (Fig. 1a). After the discovery of neural stem cells in the embryo, the first isolation of stem-like cells from adult brain^{6,7} began yet another chapter of neuroscience. Adult neural stem cells have now been found in the two principal adult neurogenic regions, the hippocampus and the subventricular zone (SVZ), and in some non-neurogenic

regions, including spinal cord⁸⁻¹⁰ (Fig. 1a). These pioneering studies provided a cellular mechanism for adult neurogenesis, which was well-established in birds and becoming accepted in mammals, and raised the possibility that the most intractable of tissues — the CNS — might have regenerative powers.

Markers that define CNS stem cells are only now being developed¹¹⁻¹⁴. Hence, they are usually identified retrospectively on the basis of their behaviour after isolation. In adherent cultures, CNS stem cells produce large clones containing neurons, glia and more stem cells; they can also be cultured as floating, multicellular neurospheres⁸⁻¹⁰. PNS neural crest stem cells express the low-affinity neurotrophin receptor p75 (ref. 5), and grow as adherent clones containing peripheral neurons and glia, smooth muscle cells and more stem cells¹⁵.

This review summarizes what we currently know about stem cells in the developing nervous system, and evaluates the idea that embryonic neural stem cells are heterogeneous and restricted. Studies that indicate broad plasticity of adult

Figure 1 The location of neural stem cells.

a. The principal regions of the embryonic and adult nervous system from which neural stem cells have been isolated^{9,10,63-65}. **b.** Three models describing stem cells in the vertebrate neural plate. All neural plate cells are stem cells (left), or stem cells are a minor population that is evenly distributed (middle) or located in particular regions such as the midline and lateral edges (right). Factors such as bone morphogenetic proteins (BMPs), Noggin, retinoids, Sonic hedgehog and fibroblast growth factors (FGFs), which provide anterior-posterior (A-P) and dorsal-ventral (D-V) patterning information, may regionalize stem cells²⁴.

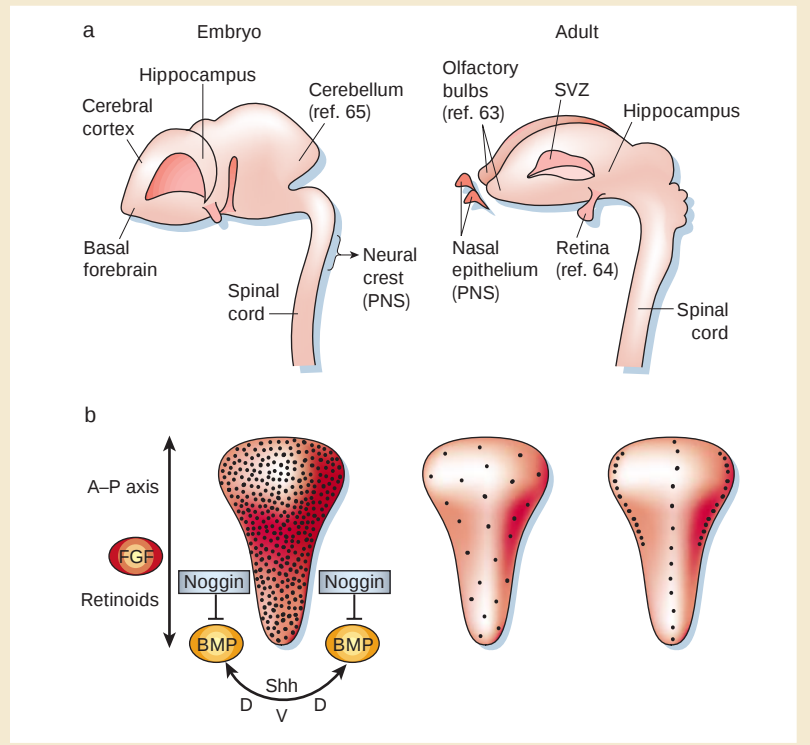
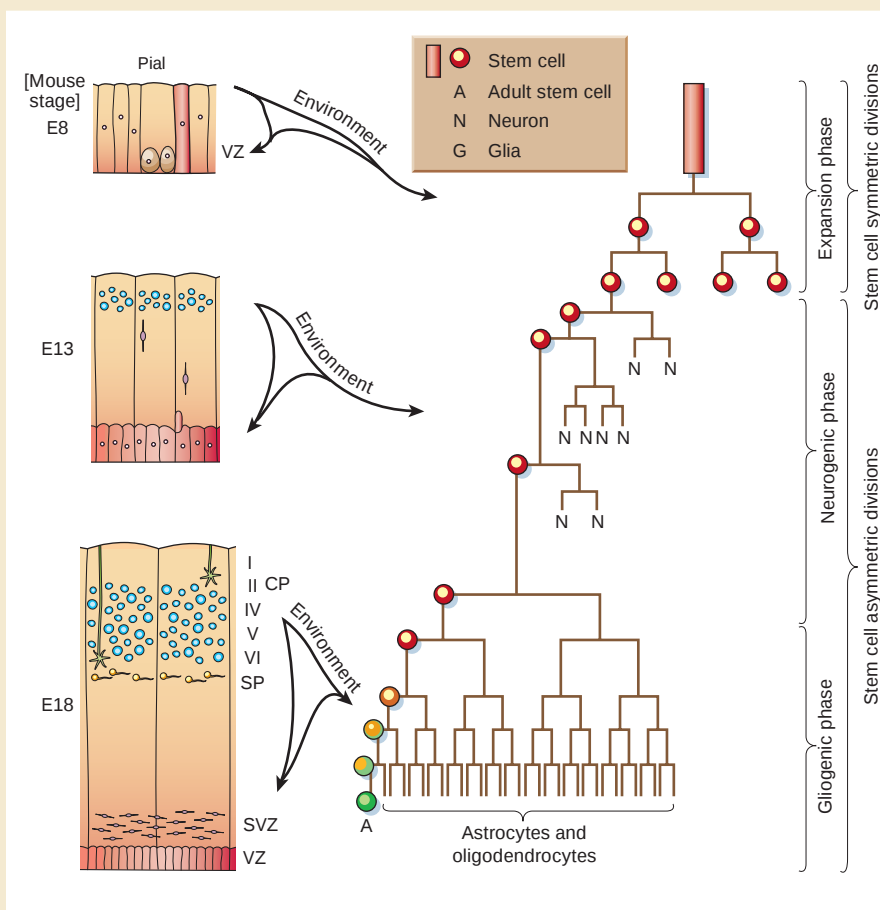


Figure 2 The development of stem cells in the mammalian CNS. Alignment of *in vivo* developmental events with *in vitro* behaviour of stem cells derived from embryonic forebrain has led to the following concept of how neural stem cells change over time. Early neuroepithelial cells are columnar, touching ventricle and pial surfaces during the cell cycle. At mid-gestation, young neurons have migrated above the germinal ventricular zone (VZ), radial glia continue to contact both ventricle and pia, guiding neuronal migration, and a second germinal zone arises, the subventricular zone (SVZ). By postnatal ages, radial glia have transformed into astrocytes and the VZ also disappears, but the SVZ remains into adulthood in some areas. Stem cells present in the early cerebral cortical neuroepithelium divide symmetrically first, and then asymmetrically to generate differentiated progeny. Neurons are produced first, and migrate along radial glia up towards the pial surface where they settle in the subplate (SP) and the cortical plate (CP). After the neurogenic period, the stem cell makes glial progenitor cells that proliferate largely in the SVZ. By birth, the stem cell has developed and has different characteristics, such as responses to growth factors, to those of the original embryonic stem cell. Stem cell development might be driven by a combination of intrinsic temporal programmes and extracellular signals from the changing environment of the developing brain.



stem cells are also discussed. By examining these two different aspects of stem cell research, future directions of exploration are highlighted that might help explain this apparent dichotomy.

Stem cells at the beginning of the nervous system

Many fundamental questions regarding specification of early neural stem cells remain unanswered. When the neural plate first emerges, does it consist solely of stem cells or does it include both stem cells and restricted progenitor types (Fig. 1)? Analysis of adherent clone production suggests stem cells are prevalent at early stages. In spinal neural tube from embryonic day 8 (E8) rat, over 50% of the viable cells at 24 hours are stem cells^{16,17}. In telencephalon from E10 mouse, estimates of stem cells range from 5 to 20% (refs 4, 18, 19). Most of the premigratory neural crest consists of stem cells²⁰.

But the frequency of stem cells declines rapidly, diluted by the production of restricted progenitors and differentiated cells; for example, in spinal cord it drops to 10% at E12 and 1% at postnatal day 1 (P1)^{16,17}. Notably, stem cells seem to be much rarer when neurosphere production is used as the assay: only 0.3% of E8.5 mouse anterior neural plate cells make neurospheres²¹. Perhaps neurosphere-generating cells are a subpopulation of early stem cells, or perhaps stem cells are more prevalent in spinal cord than anterior regions at this age.

How are neural stem cells initially specified? The stem cell could be the default state or, alternatively, stem cells might be induced. Pluripotent embryonic stem cells can produce a primitive type of neural stem cell when grown in isolation, but only 0.2% of embryonic stem cells generate neurospheres²². Although this study suggests that there is a default pathway for acquiring the stem cell state, the low frequency indicates that it may be normally enhanced by inductive mechanisms.

If stem cells are, or rapidly become, a subset of early neural progenitor cells *in vivo*, how are they distributed (Fig. 1)? Without

specific markers, important questions regarding stem cell frequency and location *in vivo* remain open.

The early, widespread presence of stem cells in the embryonic nervous system raises another important issue of their role in development. Given their prolific, multipotent nature *in vitro*, they are likely to be principal progenitors *in vivo*, but this remains to be shown directly. For example, it is possible that early restricted neuroblasts rather than stem cells might generate the preponderance of neurons. Clonal studies suggest that most glia, both astrocytes and oligodendrocytes, originate from stem cells^{9,18,19}, signifying their importance for gliogenesis. In fact, as described below, there may be an intimate association between glia and the neural stem cell state.

Neural stem cells acquire positional information

Patterning of the body axis occurs through signalling systems that impart positional information. For example, gradients of signalling molecules can regionally specify a population of progenitor cells if the cells respond differently to different concentrations of the signal²³. In the nervous system, the salient patterning in anterior–posterior and dorsal–ventral axes occurs early, concomitantly with neural induction (Fig. 1)²⁴.

Do both stem cells and restricted progenitor cells exhibit regionalization, or do stem cells remain unspecified, maintaining their plasticity? In *Drosophila* and grasshopper, each stem cell-like neuroblast has a unique identity based on its position in the neuroectoderm²⁵. In vertebrates, this question is just beginning to be explored. Neurospheres generated from different CNS regions express region-appropriate markers, indicating that the original stem cells were indeed regionally specified²⁶. Regulatory sequences control region-specific expression of the transcription factor Sox2, so that expression is seen in telencephalic but not spinal cord stem cells²⁷.

Table 1 Summary of transplantation studies assessing stem and progenitor cell behaviour

Donor	Host site	CNS incorporation	Differentiation	References
Embryonic stem cells				
Mouse ES cells or ES cell-derived neurospheres	Blastocyst	Incorporation into all embryonic tissues	Both glutamate- and GABA-mediated neurons in cortex	50,51
Mouse ES cell-derived neural precursors	E16–18 rat LV	Widespread incorporation	Neurons, oligodendrocytes and astrocytes	52
Embryonic neural cells				
E9.5 and E14.5 forebrain-derived neurospheres	Blastocyst or morula	No aggregation		22
EGF-generated neurosphere cells from E14 mouse fore- and midbrain	E15 rat LV	Fore- and midbrain structures	Astrocytes	53
Fetal human brain-derived neurospheres or primary cells	E17–18 rat or P0 mouse LV	Widespread incorporation in the brain	Neurons (projection), oligodendrocytes and astrocytes	54,55
E10.5 mid/hindbrain	E13.5 MGE	Dispersed into forebrain	Site-specific neuronal differentiation	31
E12–14 mouse forebrain	E15 to P1 rat LV	Into forebrain and midbrain; incorporation and migration reduced as host age increases	Site-specific neuronal differentiation	56,57
E13.5 mid/hindbrain	E13.5 LGE or MGE	No integration into forebrain		31
E36 ferret cortex (making layer 4)	E30 ferret (making layer 6)	Neocortex and hippocampus	Layers 2–5; no transplanted cells in layer 6	29
	P2 (making 2/3)	Neocortex and hippocampus	Layer 2/3 neurons	
Early postnatal				
Postnatal mouse SVZ cells	E15 mouse LV	Septum, thalamus, hypothalamus and inferior colliculus; not cortex or hippocampus	No principal projection neurons	32
Neonatal mouse SVZ neurospheres	Chick embryo	Migrate in isolation	Neural crest derivatives	46
Neonatal mouse SVZ cells	Chick embryo	Chain migration in the neural crest pathway	Progenitors and neurons; not neural crest phenotypes	46
	Neonatal striatum	Migration	Neuronal precursors and olfactory bulb neurons	58
P0–5 mouse SVZ cells	Adult striatum	Some migration	Olfactory bulb interneurons	59
Adult				
Mouse forebrain neurospheres	Blastocyst or morula	No aggregation		22
	Mouse morula or chick embryo	Rare neural chimeras	Cell types not reported	60
Mouse SVZ neurospheres	Chick embryo	Migrate as isolated cells	Phenotypes of neural crest derivatives	46
Mouse SVZ cells	Chick embryo	Cells die		46
	Adult mouse striatum	Minimum migration	Some neurons, mostly astrocytes in striatum	61
	Adult olfactory bulb	Extensive migration	Olfactory bulb neurons	61
Cultured rat hippocampal progenitors	Adult rat hippocampus	Migrate	Neurons in granule cell layer of dentate gyrus	42, 62
	Adult rat rostral migratory stream	Migrate to olfactory bulb	Neuroblasts and olfactory neurons	42
	Adult rat cerebellum	Incorporation	No neurons	42
FGF-responsive rat spinal cord progenitors	Adult hippocampus	Broad dispersion	Hippocampal granule neurons	41
	Adult spinal cord	Broad dispersion	No neurons	41

These studies indicate stage-dependent restrictions in the potential of donor progenitor populations (which include some neural stem cells). They also emphasize the impact of interaction between implanted cells and the host environment. Different treatments of stem and progenitor cells *in vitro* can significantly alter their behaviour after implantation. EGF, epidermal growth factor; ES, embryonic stem; FGF, fibroblast growth factor. LGE/MGE, lateral/medial ganglionic eminence; LV, lateral ventricle.

In addition, stem cells isolated from different neural regions generate region-appropriate progeny. Spinal cord stem cells generate spinal cord progeny¹⁷. Basal forebrain stem cells cultured at clonal density generate significantly more GABA (γ -amino butyric acid)-containing neurons, which are characteristic of basal regions, than dorsal stem cells cultured under identical conditions¹⁹. PNS neural crest stem cells, which arise at the lateral edges of the neural plate, express distinct genes and generate progeny distinct from those of CNS stem cells^{15,28}. Hence, vertebrate stem cells seem to be positionally specified.

Neural stem cells acquire temporal information

Different neural cell types arise in a precise temporal order that is characteristic for a particular region and species. In general, CNS and PNS neurons arise before glia, and specific types of each cell have specific birthdates. Timing seems to be encoded in progenitor cells, so that besides positional information they have 'temporal information', which is seen as stage-dependent changes in progenitor cells (Table 1). Thus, late-embryonic ferret cortical progenitor cells

cannot make cells appropriate for younger stages when transplanted into early cerebral cortex²⁹. Rat cortical progenitors become restricted in their ability to generate limbic system-associated membrane protein (LAMP)-positive limbic cortical progeny after E14 (ref. 30). Mid/hindbrain progenitor cells are unable to generate telencephalic phenotypes after E13.5 in mouse³¹. SVZ cells can no longer make projection neurons by birth^{32,33}, and retinal progenitor cells seem to be similarly restricted temporally^{34,35}.

Stem cells are a minor component of the progenitor population in these studies, but experiments indicate that they also exhibit stage-dependent changes in potential. Some early neural tube cells produce both CNS and PNS stem cells, suggesting that there is a common progenitor for two separate stem cell lineages with more restricted potentials²⁸. Each of these lineages also changes over time. CNS stem cells undergo repeated asymmetric cell divisions, first producing neurons then glia¹⁸, thus reproducing the normal neuron–glia order (Fig. 2). Moreover, they have an active role in this process by altering their intrinsic properties. Thus, stem cells from earlier stage cortices

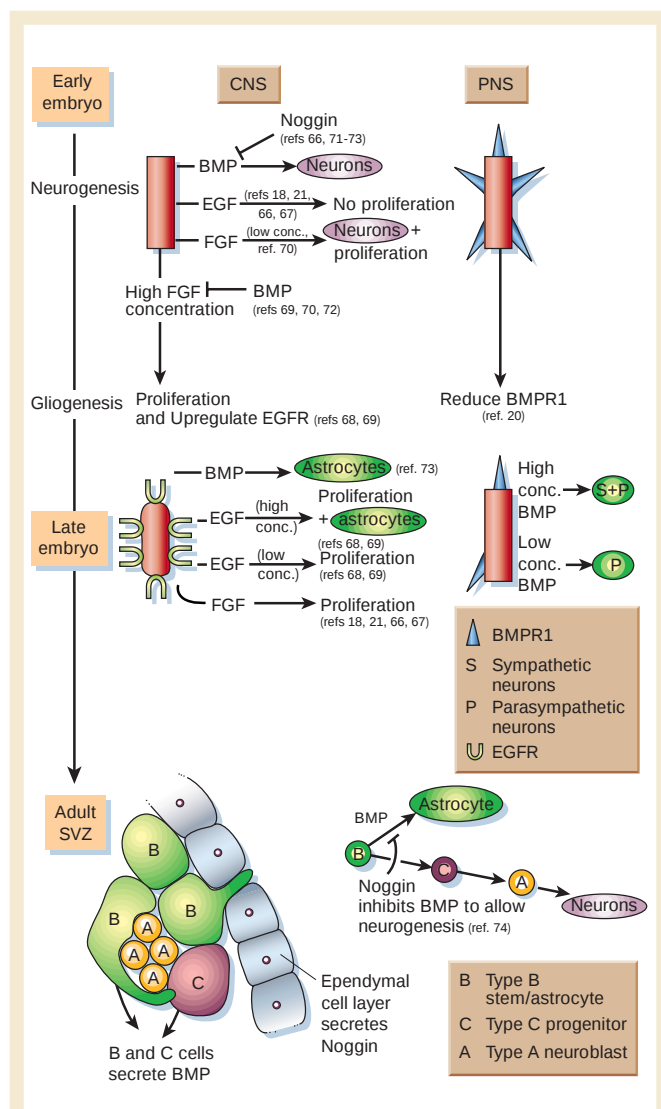


Figure 3 Stem cells alter their responses to growth factors over time. Central and peripheral nervous system (neural crest) stem cells extracted from different ages show differing responses to application of the same growth factor. Changes in levels of growth factor receptors have been observed. In addition, stem cells exhibit different responses to growth factors applied at different concentrations. BMP, bone morphogenetic protein; EGF, epidermal growth factor; EGFR, EGF receptor; FGF, fibroblast growth factor.

produce more neurons and have a lower tendency to produce glia than those from later stages.

Temporal specification of stem cell populations also occurs in the PNS. Mouse neural crest stem cells isolated from the rat neural tube have been compared with later stem cells isolated prospectively from E14 sciatic nerve after transplantation into different sites of the chick embryo²⁰. Early neural crest stem cells generate significantly more neurons than later stage cells: like CNS stem cells, their neurogenic capacity declines with stage. Furthermore, the range of neurons generated by late neural crest stem cells is more restricted. Early transplants that are highly enriched for neural crest stem cells but contain some restricted sensory progenitor cells can generate dorsal root ganglion (DRG) sensory neurons, and adrenergic and cholinergic autonomic neurons. By contrast, older-stage neural crest stem cells made no DRG neurons, only rare adrenergic sympathetic neurons, but they could generate cholinergic autonomic neurons.

Stem cells also undergo phenotypic changes as germinal zones develop (Fig. 2). It has been suggested recently that some radial glia

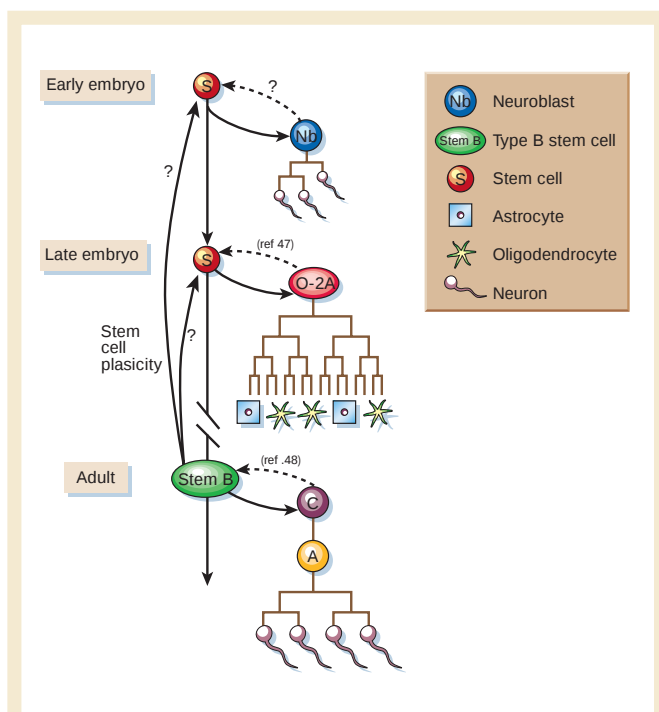


Figure 4 Can the stem cell lineage be reversed? Evidence indicates that neural stem cell progeny can reacquire stem cell properties. An important issue is whether adult stem cells can revert to an earlier embryonic state, with a wider potential.

present at mid-gestation might be stem cells (reviewed in ref. 36). This idea is appealing, given that radial glia are thought to be neurogenic precursors in adult canary brains, and that astrocytes, the lineal descendants of radial glia, have stem cell properties in the adult mammalian SVZ³⁷.

Developmental changes in stem cells — for example, in their potential and phenotype — are accomplished by, and perhaps driven by, changes in their growth factor responsiveness (Fig. 3; and see accompanying reviews by Spradling and colleagues, pages 98–104, and Weissman and colleagues, pages 105–111). Thus signalling molecules, such as fibroblast growth factors, bone morphogenetic proteins and Noggin, can influence neural stem cells from neural induction through adulthood, but their responses to these factors vary with stage.

The mechanisms underlying temporal changes in neural stem cells are not understood. The lineage trees of mouse CNS stem cells are remarkably similar to those of invertebrates³⁸. *Drosophila* CNS neuroblasts express sequentially the transcription factors Hunchback, Krüppel, POU-domain genes 1 and 2, Castor and grainyhead, which specify the production of different neurons at different times^{39,40}. The expression sequence may be driven by a cell-intrinsic clock⁴⁰. Perhaps similar intrinsic timing mechanisms, combined with environmental input, temporally specify mammalian CNS stem cells. Growth factor concentrations vary during development and neural stem cells respond differently to different concentrations of the same molecule (Fig. 3). Hence, it is tempting to speculate that, just as positional information can be imparted by spatial gradients of signals²³, temporal information might be imparted by temporal gradients of signalling molecules.

Examining the plasticity of adult stem cells

If stem cells undergo developmental changes, adult stem cells are likely to differ from those in the embryo and to be regionally and temporally restricted.

One way to examine the plasticity of adult neural stem cells is to transplant them directly into the developing embryo and examine

Box 1

Potential therapeutic uses of stem cells to repair the nervous system

Stem cells are under active consideration as a source of donor tissues for neuronal cell therapy⁷⁵.

Parkinson's disease. The requirement is to generate cells that synthesize and release dopamine for implantation into the dopamine-depleted striatum. For this therapy to be effective, it is unknown whether these cells must also mature into projection neurons with synaptic host connections — a process that is required for optimal effects of embryonic nigral grafts.

Huntington's disease. If we can control differentiation into mature neuronal phenotypes then many other diseases that involve loss of specific neuronal types, such as the striatal medium spiny projection neurons lost in Huntington's disease, might be suitable for transplantation of neurons expanded from stem cell sources.

Spinal cord injury. Stem cells may be able to repopulate the site of injury, provide a substrate for axon growth across the transection, and block syrinx progression. Each of these effects has been found with primary embryonic cells; however, studies using stem or immortalized precursors are still preliminary.

Stroke. Stem cells and immortalized precursors may be able to migrate through the central nervous system and repopulate sites of ischaemia. In spite of rather limited evidence from animal studies, clinical trials of this strategy are already underway.

Multiple sclerosis. Oligodendrocyte lineages are better characterized than neuron lineages, and oligodendrocyte precursors can differentiate and provide a functional remyelination of axons after focal experimental demyelination. For application in multiple sclerosis, the main problem is how to stimulate the migration of such cells to diverse sites of demyelination that occur sporadically in the human disease. Notwithstanding the potential applications of oligodendrocyte lineages in several diseases, many key technical problems remain to be resolved.

Sources. If we select early embryonic stem cells, then the number of transformations and the complexity of signals required to achieve a specific differentiated phenotype may prove prohibitive; instead, it may be easier to control the phenotypic differentiation of developing neuronal precursors, but that might in turn limit their capacity for expansion. Adult-derived cells may circumvent both ethical and immunological constraints, but their plasticity for expansion and differentiation remains to be established. Cross-lineage transformation offers a new prospect for a more flexible source, in particular to derive autografts from patients themselves. A further advance is that precursor cells are less immunogenic than primary embryonic neurons in xenografts, highlighting a way to overcome one of the main difficulties of transplantation from non-human donors.

Expansion. Neuronal stem cells from species other than mice seem to senesce on repeated passage, with only limited potential for expansion. Conversely, if not fully differentiated at the time of implantation, there is always the possibility of tumour formation — a problem that is still not resolved for either embryonic stem cells or immortalized precursors.

Differentiation. The biggest single problem still to be solved is how to direct and control the differentiation of specific target phenotypes required for replacement and repair in each disease. Selection of appropriate starting cells by embryonic regional dissection and stage of development, as well as diverse parameters of *in vitro* manipulation, are likely to be crucial factors in directing appropriate phenotypic differentiation. Failure can lead not only to a lack of benefit but also to significant side-effects from proliferation of non-neuronal phenotypes.

All these problems can be solved, but it will require more than a single breakthrough to transform the potential attributed to stem cells into a realistic clinical strategy for cellular repair.

what neural cell types they produce; this still remains to be done. Emerging methods for the prospective isolation of adult CNS stem cells^{11–14} should facilitate this important experiment. Experiments using culture-expanded adult CNS stem cells indicate that there is some plasticity in adult environments (Table 1). For example, adult spinal-cord-derived stem cells, which do not normally make neurons, can make interneurons if injected into the adult hippocampus⁴¹, and adult hippocampal-derived stem cells can make olfactory interneurons after transplantation to the SVZ⁴². As yet, however, there is no direct evidence that adult-derived stem cells can make the types of projection neuron that are normally generated in the embryo.

Given the vital gaps in our understanding of adult neural stem cells, we cannot yet conclude that they are highly plastic. Evidence that they can generate different somatic cell types is limited, and may be restricted to rare events or rare cells^{43–45}. In considering the two different ideas raised at the beginning of this review, there may be in fact no dichotomy. Most neural stem cells might be regionally and temporally specified. There may also be rare stem cells present in the nervous system, perhaps not even of neural origin, that have greater plasticity, at least in terms of producing diverse somatic cell types⁴³.

Reversing the stem cell lineage

If stem cells are restricted, can these restrictions be reversed (Fig. 4)? For example, can an adult stem cell re-acquire the ability to generate cell types normally made in the embryo? A study has indicated that culturing SVZ stem cells as neurospheres expands their potential and allows them to generate PNS progeny after injection into chick neural crest pathways⁴⁶. Furthermore, descendants of neural stem cells may

be able to revert. Optic nerve O2A progenitor cells, which normally produce solely glia, can be converted *in vitro* into multipotent, neurosphere-generating stem cells⁴⁷. In the adult SVZ, both type B stem cells and their progeny, type C progenitor cells, can make neurospheres *in vitro*. Moreover, type C cells are stimulated to convert to stem cells by epidermal growth factor⁴⁸. In the postnatal chick retina after damage, Muller radial glia, which are maintained throughout life, can re-enter the cell cycle, re-express retinal progenitor cell markers and generate new neurons and glia⁴⁹.

If environmental factors can enhance the acquisition of neural stem cell fates, or increase the plasticity of stem cells, this may be of enormous benefit therapeutically, as indicated in Box 1.

Future studies

As regions of the embryo are patterned and development unfolds, neural stem cells may be an essential mediator of developmental signals, acquiring a changing repertoire of gene expression, morphology and behaviour. Despite differences in the properties of stem cells isolated from different regions and at different times, they still self-renew. Self-renewal can therefore be considered as the propagation of stem cells, rather than the production of exactly the same type of cell.

It will be important to examine how developmental signals, both spatial and temporal, specify changes in neural stem cells. Markers for neural stem cells will allow their selection from different stages and regions to examine their potential after transplantation into the embryo or adult, and a comparison of their gene expression. Such explorations will help identify essential mediators of stem cell self-renewal, and genes that determine production of different types of progeny. Markers will also help solve the tantalizing issue of which cells *in vivo* are stem cells.

Although research to identify adult sources of highly plastic stem cells for therapeutic use will continue, it seems likely that most neural-generating stem cells might be specified during development. In that case, we must explore this diversity to understand how different neural cells are, and can be, made. □

1. Temple, S. Division and differentiation of isolated CNS blast cells in microculture. *Nature* **340**, 471–473 (1989).
2. Cattaneo, E. & McKay, R. Proliferation and differentiation of neuronal stem cells regulated by nerve growth factor. *Nature* **347**, 762–765 (1990).
3. Reynolds, B. A., Tetzlaff, W. & Weiss, S. A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes. *J. Neurosci.* **12**, 4565–4574 (1992).
4. Kilpatrick, T. J. & Bartlett, P. F. Cloning and growth of multipotential neural precursors: requirements for proliferation and differentiation. *Neuron* **10**, 255–265 (1993).
5. Stemple, D. L. & Anderson, D. J. Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell* **71**, 973–985 (1992).
6. Reynolds, B. A. & Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **255**, 1707–1710 (1992).
7. Lois, C. & Alvarez-Buylla, A. Proliferating subventricular zone cells in the adult mammalian forebrain can differentiate into neurons and glia. *Proc. Natl Acad. Sci. USA* **90**, 2074–2077 (1993).
8. McKay, R. Stem cells in the central nervous system. *Science* **276**, 66–71 (1997).
9. Rao, M. S. Multipotent and restricted precursors in the central nervous system. *Anat. Rec.* **257**, 137–148 (1999).
10. Gage, F. H. Mammalian neural stem cells. *Science* **287**, 1433–1438 (2000).
11. Capela, A. & Temple, S. A putative surface marker for adult mouse neural stem cells. *Soc. Neurosci. Abstr.* **24.19** <http://sfn.scholarone.com/itin2000/> (2000).
12. Kawaguchi, A. *et al.* Nestin-EGFP transgenic mice: visualization of the self-renewal and multipotency of CNS stem cells. *Mol. Cell. Neurosci.* **17**, 259–273 (2001).
13. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl Acad. Sci. USA* **97**, 14720–14725 (2000).
14. Rietze, R. L. *et al.* Purification of a pluripotent neural stem cell from the adult mouse brain. *Nature* **412**, 736–739 (2001).
15. Anderson, D. J. *et al.* Cell lineage determination and the control of neuronal identity in the neural crest. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 493–504 (1997).
16. Kalyani, A., Hobson, K. & Rao, M. S. Neuroepithelial stem cells from the embryonic spinal cord: isolation, characterization, and clonal analysis. *Dev. Biol.* **186**, 202–223 (1997).
17. Kalyani, A. J., Piper, D., Mujtaba, T., Lucero, M. T. & Rao, M. S. Spinal cord neuronal precursors generate multiple neuronal phenotypes in culture. *J. Neurosci.* **18**, 7856–7868 (1998).
18. Qian, X. *et al.* Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron* **28**, 69–80 (2000).
19. He, W., Ingraham, C., Rising, L., Goderie, S. & Temple, S. Multipotent stem cells from the mouse basal forebrain contribute GABAergic neurons and oligodendrocytes to the cerebral cortex during embryogenesis. *J. Neurosci.* (in the press).
20. White, P. M. *et al.* Neural crest stem cells undergo cell-intrinsic developmental changes in sensitivity to instructive differentiation signals. *Neuron* **29**, 57–71 (2001).
21. Tropepe, V. *et al.* Distinct neural stem cells proliferate in response to EGF and FGF in the developing mouse telencephalon. *Dev. Biol.* **208**, 166–188 (1999).
22. Tropepe, V. *et al.* Direct neural fate specification from embryonic stem cells. A primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron* **30**, 65–78 (2001).
23. Wolpert, L. Positional information and pattern formation in development. *Dev. Genet.* **15**, 485–490 (1994).
24. Altmann, C. R. & Brivanlou, A. H. Neural patterning in the vertebrate embryo. *Int. Rev. Cytol.* **203**, 447–482 (2001).
25. Skeath, J. B. At the nexus between pattern formation and cell-type specification: the generation of individual neuroblast fates in the *Drosophila* embryonic central nervous system. *BioEssays* **21**, 922–931 (1999).
26. Tropepe, V., Hitoshi, S., Ekker, M. & van der Kooy, D. Neural stem cells and their progeny express region-specific genes in the developing CNS, but this expression is not irreversible and can be altered by local inductive cues. *Soc. Neurosci. Abstr.* **23.1** <http://sfn.scholarone.com/itin2000/> (2000).
27. Zappone, M. V. *et al.* Sox2 regulatory sequences direct expression of a β -geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382 (2000).
28. Mujtaba, T., Mayer-Proschel, M. & Rao, M. S. A common neural progenitor for the CNS and PNS. *Dev. Biol.* **200**, 1–15 (1998).
29. Desai, A. R. & McConnell, S. K. Progressive restriction in fate potential by neural progenitors during cerebral cortical development. *Development* **127**, 2863–2872 (2000).
30. Barbe, M. F. & Levitt, P. The early commitment of fetal neurons to the limbic cortex. *J. Neurosci.* **11**, 5159–5333 (1991).
31. Olson, M., Campbell, K. & Turnbull, D. H. Specification of mouse telencephalic and mid-hindbrain progenitors following heterotopic ultrasound-guided embryonic transplantation. *Neuron* **19**, 761–772 (1997).
32. Lim, D. A., Fishell, G. J. & Alvarez-Buylla, A. Postnatal mouse subventricular zone neuronal precursors can migrate and differentiate within multiple levels of the developing neuraxis. *Proc. Natl Acad. Sci. USA* **94**, 14832–14836 (1997).
33. Alvarez-Buylla, A., Herrera, D. G. & Wichterle, H. The subventricular zone: source of neuronal precursors for brain repair. *Prog. Brain Res.* **127**, 1–11 (2000).
34. Reh, T. A. & Levine, E. M. Multipotential stem cells and progenitors in the vertebrate retina. *J. Neurobiol.* **36**, 206–220 (1998).
35. Livesey, F. J. & Cepko, C. L. Vertebrate neural cell-fate determination: lessons from the retina. *Nature Rev. Neurosci.* **2**, 109–118 (2001).
36. Alvarez-Buylla, A., Garcia-Verdugo, J. M. & Tramontin, A. D. A unified hypothesis on the lineage of neural stem cells. *Nature Rev. Neurosci.* **2**, 287–293 (2001).
37. Doetsch, F., Caille, L., Lim, D. A., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–716 (1999).
38. Shen, Q., Qian, X., Capela, A. & Temple, S. Stem cells in the embryonic cerebral cortex: their role in histogenesis and patterning. *J. Neurobiol.* **36**, 162–174 (1998).
39. Brody, T. & Odenwald, W. F. Programmed transformations in neuroblast gene expression during *Drosophila* CNS lineage development. *Dev. Biol.* **226**, 34–44 (2000).
40. Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511–521 (2001).
41. Shihabuddin, L. S., Horner, P. J., Ray, J. & Gage, F. H. Adult spinal cord stem cells generate neurons after transplantation in the adult dentate gyrus. *J. Neurosci.* **20**, 8727–8735 (2000).
42. Suhonen, J. O., Peterson, D. A., Ray, J. & Gage, F. H. Differentiation of adult hippocampus-derived progenitors into olfactory neurons *in vivo*. *Nature* **383**, 624–627 (1996).
43. Weissman, I. L. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100**, 157–168 (2000).
44. Anderson, D. J. Stem cells and pattern formation in the nervous system. The possible versus the actual. *Neuron* **30**, 19–35 (2001).
45. Temple, S. Stem cell plasticity—building the brain of our dreams. *Nature Neurosci. Rev.* **2**, 513–520 (2001).
46. Durbec, P. & Rougon, G. Transplantation of mammalian olfactory progenitors into chick hosts reveals migration and differentiation potentials dependent on cell commitment. *Mol. Cell. Neurosci.* **17**, 561–576 (2001).
47. Kondo, T. & Raff, M. Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. *Science* **289**, 1754–1757 (2000).
48. Doetsch, F. K., Caille, L., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. EGF induces conversion of transit amplifying neurogenic precursors into multipotential invasive cells in the adult brain. *Soc. Neurosci. Abstr.* **24.19** <http://sfn.scholarone.com/itin2001/> (2001).
49. Fischer, A. J. & Reh, T. A. Muller glia are a potential source of neural regeneration in the postnatal chicken retina. *Nature Neurosci.* **4**, 247–252 (2001).
50. Goldowitz, D. Cell allocation in mammalian CNS formation: evidence from murine interspecies aggregation chimeras. *Neuron* **3**, 705–713 (1989).
51. Tan, S. S. *et al.* Separate progenitors for radial and tangential cell dispersion during development of the cerebral neocortex. *Neuron* **21**, 295–304 (1998).
52. Brustle, O. *et al.* *In vitro*-generated neural precursors participate in mammalian brain development. *Proc. Natl Acad. Sci. USA* **94**, 14809–14814 (1997).
53. Winkler, C. *et al.* Incorporation and glial differentiation of mouse EGF-responsive neural progenitor cells after transplantation into the embryonic rat brain. *Mol. Cell. Neurosci.* **11**, 99–116 (1998).
54. Brustle, O. *et al.* Chimeric brains generated by intraventricular transplantation of fetal human brain cells into embryonic rats. *Nature Biotechnol.* **16**, 1040–1044 (1998).
55. Flax, J. D. *et al.* Engraftable human neural stem cells respond to developmental cues, replace neurons, and express foreign genes. *Nature Biotechnol.* **16**, 1033–1039 (1998).
56. Brustle, O., Maskos, U. & McKay, R. D. Host-guided migration allows targeted introduction of neurons into the embryonic brain. *Neuron* **15**, 1275–1285 (1995).
57. Campbell, K., Olsson, M. & Bjorklund, A. Regional incorporation and site-specific differentiation of striatal precursors transplanted to the embryonic forebrain ventricle. *Neuron* **15**, 1259–1273 (1995).
58. Betarbet, R., Zigova, T., Bakay, R. A. & Luskin, M. B. Migration patterns of neonatal subventricular zone progenitor cells transplanted into the neonatal striatum. *Cell Transplant.* **5**, 165–178 (1996).
59. Lois, C. & Alvarez-Buylla, A. Long-distance neuronal migration in the adult mammalian brain. *Science* **264**, 1145–1148 (1994).
60. Clarke, D. L. *et al.* Generalized potential of adult neural stem cells. *Science* **288**, 1660–1663 (2000).
61. Herrera, D. G., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Adult-derived neural precursors transplanted into multiple regions in the adult brain. *Ann. Neurol.* **46**, 867–877 (1999).
62. Gage, F. H. *et al.* Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. *Proc. Natl Acad. Sci. USA* **92**, 11879–11883 (1995).
63. Pagano, S. F. *et al.* Isolation and characterization of neural stem cells from the adult human olfactory bulb. *Stem Cells* **18**, 295–300 (2000).
64. Tropepe, V. *et al.* Retinal stem cells in the adult mammalian eye. *Science* **287**, 2032–2036 (2000).
65. Laywell, E. D., Rakic, P., Kukekov, V. G., Holland, E. C. & Steindler, D. A. Identification of a multipotent astrocytic stem cell in the immature and adult mouse brain. *Proc. Natl Acad. Sci. USA* **97**, 13883–13888 (2000).
66. Zhu, G., Mehler, M. F., Mabie, P. C. & Kessler, J. A. Developmental changes in progenitor cell responsiveness to cytokines. *J. Neurosci. Res.* **56**, 131–145 (1999).
67. Gritti, A. *et al.* Epidermal and fibroblast growth factors behave as mitogenic regulators for a single multipotent stem cell-like population from the subventricular region of the adult mouse forebrain. *J. Neurosci.* **19**, 3287–3297 (1999).
68. Burrows, R. C., Wancio, D., Levitt, P. & Lillien, L. Response diversity and the timing of progenitor cell maturation are regulated by developmental changes in EGFR expression in the cortex. *Neuron* **19**, 251–267 (1997).
69. Lillien, L. & Raphael, H. BMP and FGF regulate the development of EGF-responsive neural progenitor cells. *Development* **127**, 4993–5005 (2000).
70. Qian, X., Davis, A. A., Goderie, S. K. & Temple, S. FGF2 concentration regulates the generation of neurons and glia from multipotent cortical stem cells. *Neuron* **18**, 81–93 (1997).
71. Li, W., Cogswell, C. A. & LoTurco, J. J. Neuronal differentiation of precursors in the neocortical ventricular zone is triggered by BMP. *J. Neurosci.* **18**, 8853–8862 (1998).
72. Li, W. & LoTurco, J. J. Noggin is a negative regulator of neuronal differentiation in developing neocortex. *Dev. Neurosci.* **22**, 68–73 (2000).
73. Mehler, M. F., Mabie, P. C., Zhu, G., Gokhan, S. & Kessler, J. A. Developmental changes in progenitor cell responsiveness to bone morphogenetic proteins differentially modulate progressive CNS lineage fate. *Dev. Neurosci.* **22**, 74–85 (2000).
74. Lim, D. A. *et al.* Noggin antagonizes BMP signaling to create a niche for adult neurogenesis. *Neuron* **28**, 713–726 (2000).
75. Kirschstein, R. & Skirboll, L. R. *Stem Cells: Scientific Progress and Future Research Directions* (NIH, Bethesda, 2001).

Acknowledgements

I thank Q. Shen for her help with Table 1 and for invaluable discussions; K. Kirchofer for preparing the manuscript; and S. Dunnett for summarizing the potential therapeutic applications of stem cells in Box 1.

Stem cells in tissue engineering

Paolo Bianco* & Pamela Gehron Robey†

*Dipartimento di Medicina Sperimentale e Patologia, Università 'La Sapienza', Viale Regina Elena 324, Roma 00161, Italy (e-mail: p.bianco@flashnet.it)

†Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive MSC 4320, Bethesda, Maryland, USA (e-mail: probey@dir.nidcr.nih.gov)

The concept of producing 'spare parts' of the body for replacement of damaged or lost organs lies at the core of the varied biotechnological practices referred to generally as tissue engineering. Use of postnatal stem cells has the potential to significantly alter the perspective of tissue engineering. Successful long-term restoration of continuously self-renewing tissues such as skin, for example, depends on the use of extensively self-renewing stem cells. The identification and isolation of stem cells from a number of tissues provides appropriate targets for prospective gene therapies.

Tissue engineering in a classical sense implies the use of organ-specific cells for seeding a scaffold *ex vivo*. The cell-based (but not necessarily stem cell-based) nature of tissue engineering served to define it, and to distinguish it from 'guided tissue regeneration' in which a scaffold is designed to encourage regeneration solely by cells residing at the site of its transplantation. Irrespective of actual clinical feasibility, the experimental design and potential use of tissue engineering approaches are endless, and range from the preclinical generation of cardiac valve substitutes, to *ex vivo* construction of nasal cartilages, to organ substitutes such as the urinary bladder (reviewed in ref. 1).

The list of tissues with the potential to be engineered is growing steadily. This is due in large part to recent progress in stem cell biology and recognition of the unique biological properties of stem cells, although not all stem cell-based therapies (for example, some that use neural stem cells, as summarized by Temple, pages 112–117) involve the construction of tissue either *ex vivo* or *in vivo*. Nonetheless, pilot studies in a variety of systems highlight great prospects for future stem cell-based tissue engineering (Table 1). But translation into clinical reality has been reached so far in only a few areas, and notably only in those areas in which there has been long-standing insight into stem cell biology.

Tissues that can now be engineered using stem cells comprise a diverse range from epithelial surfaces (skin, cornea and mucosal membranes) to skeletal tissues. These systems are inherently different in their rate of self-renewal and physical structure, two important determinants of any attempt to reconstruct tissues using stem cells (as discussed by Spradling and colleagues on pages 98–104). Appreciation of the inherent diversities of organ systems and cognate stem cells is essential for development of adequate strategies of intervention in specific areas.

Here we examine two applications of stem cells to tissue engineering. The first case is the regeneration of skin, which involves the structural formation of two-dimensional sheets. The second, more complex example is the formation of bone, which involves the reconstruction of three-dimensional shapes and internal architectures.

Engineering skin and other surfaces

Permanent restoration of tissues characterized by high and continuous self-renewal specifically requires extensively self-renewing stem cells. Haematopoiesis provides the best paradigm of a stem cell-dependent, steadily self-

renewing system (see review by Weissman and colleagues, pages 105–111). Among the hierarchy of haematopoietic progenitors endowed with varying capacities to self-renew, it is only the small compartment of long-term self-renewing stem cells that is necessary for permanent and complete restoration of haematopoiesis after lethal irradiation². Likewise, progenitor cells represented in a keratinocyte culture are also ranked in a hierarchy, and under clonogenic conditions form different types of colonies (holoclones, meroclones and paraclones) that vary significantly in their ability to self-renew and to generate a differentiated epidermis. Only holoclones are the product of a true stem cell, as defined by their exceptional self-replication ability (>140 doublings), which is not ascribed to either meroclones (a population of transient amplifying cells) or paraclones (a population of senescent or differentiating progenitors)³. In principle, a small but pure population of holoclone-generating cells would be all that is required for generating epidermal grafts. The recent identification of a keratinocyte stem cell marker may provide a new tool for this approach with respect to epidermis⁴.

These theoretical predictions correlate with the clinical outcomes of skin grafting. Skin autografts are produced by culturing keratinocytes to generate an epidermal sheet, and then transplanting this sheet along with a suitable dermis-like substrate (Fig. 1a)⁵. But the success of these procedures has been variable, with graft failure resulting in many cases after a promising initial engraftment. Experimental conditions can cause depletion of the holoclone-generating compartment in cell cultures, similar to the partial depletion of the epidermal stem cell compartment that can occur *in vivo* (for example, during ageing). Retention or depletion of true stem cells in a keratinocyte population *in vitro* is significantly affected by the nature of the carrier or substrate used for culturing and grafting, and could account for instances of graft failure in clinical practice⁶. The long-term success of a skin graft thus depends on appropriate replenishment of stem cells in the graft. The specific technological challenge in restoration of epithelial surfaces in turn consists of the definition of culture conditions, and of carriers that are designed specifically to maintain an adequate stem cell compartment in the engineered graft. Dependence of epithelial surfaces on relevant stem cell compartments is further highlighted by the successful reconstruction of damaged corneas using stem cells derived from the limbus in conjunction with an amniotic membrane^{7,8}.

Engineering the correct shape and physical structure of a graft is relatively simple in epidermis and other surface epithelial tissues, and indeed can be achieved *ex vivo*. This reflects the essentially two-dimensional organization of skin and other epidermal surfaces, and is in sharp contrast with systems such as the skeleton, where effective engineering involves the reconstruction of complex shapes and architectures. These directly affect proper function, are normally generated over a long period of time, and cannot be achieved *ex vivo*.

Engineering the skeleton

Skeletal stem cells (SSCs; also known as bone-marrow stromal stem cells, or mesenchymal stem cells) are found in the subset of clonogenic adherent marrow-derived cells, and are able to undergo extensive replication in culture. Upon ectopic *in vivo* transplantation in model systems, all of the main tissues found in bone as an organ (bone, cartilage, adipocytes and haematopoiesis-supporting stroma) are formed (Figs 1b and 2a,f) (ref. 9, and reviewed in refs 10,11). SSCs obtained in culture must be combined with appropriate carriers before transplantation. These provide a three-dimensional scaffold in which a vascular bed can be established, and transplanted progenitor cells can differentiate and form a bone/marrow organ. Many suitable materials are being generated and perfected, including synthetic hydroxyapatite/tricalcium phosphates (Figs 1b and 2a), and polyglycolic and polylactic acids^{12,13}. Most are effective in supporting bone regeneration, either alone or in conjunction with growth factors such as bone morphogenetic proteins. However, so far, little attention has been devoted to their compatibility with long-term maintenance of stem cell properties, or to the ultimate fate of the bone–biomaterial composite being generated at the site of transplantation.

In the simplest procedure for bone reconstruction, SSCs isolated in culture from a limited volume of bone marrow are expanded *ex vivo*, loaded onto an appropriate carrier, and locally transplanted (Fig. 1b). This procedure results in the effective repair of critical size defects, which are larger than what can be repaired by resident cells, either spontaneously or as guided by an osteoconductive device (Figs 1b and 2b,c)^{14–17}. Convincing preclinical studies adopting this approach have led to preliminary observational studies in humans¹⁸, and clinical trials are about to start in a number of centres. An alternative approach has led to the generation of fully vascularized bone flaps of the desired shape *in vivo*, prior to their use for local transplantation (Fig. 2d, e). In these experiments, SSCs loaded into appropriate carriers and transplanted into a non-skeletal site surrounding an artery and vein; here they generate a vascularized segment of bone, the size and shape of which are dictated by the carrier geometry¹⁹.

All of these applications can be categorized as surgical intervention — through transplantation of a ‘biological’ prosthetic device — for the cure of physical defects in an otherwise normal skeleton. But the existence of SSCs raises additional hopes and challenges for treatment of more complex and generalized diseases, such as crippling genetic diseases. Here, the concept of using SSCs to replace malfunctioning bone cells with normal ones meets another set of important obstacles, which further illustrate the diversity of the skeleton and its stem cells compared with other stem cell-based systems. One is the route of systemic delivery of the number of progenitors needed to produce a biological effect. Some

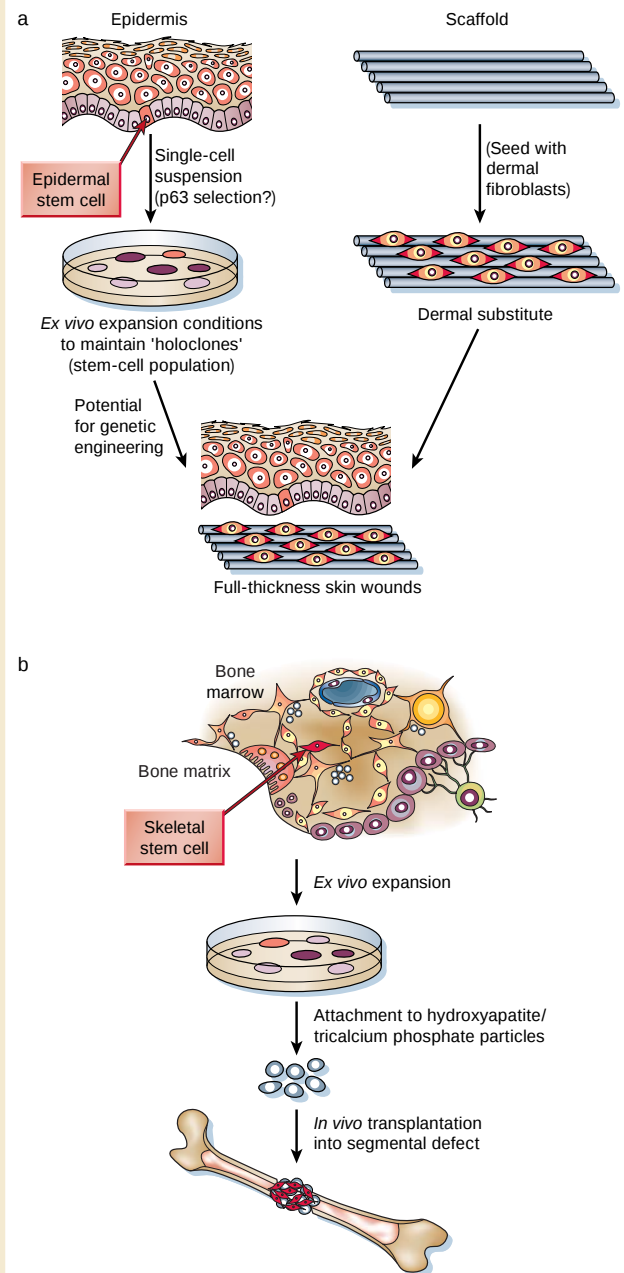


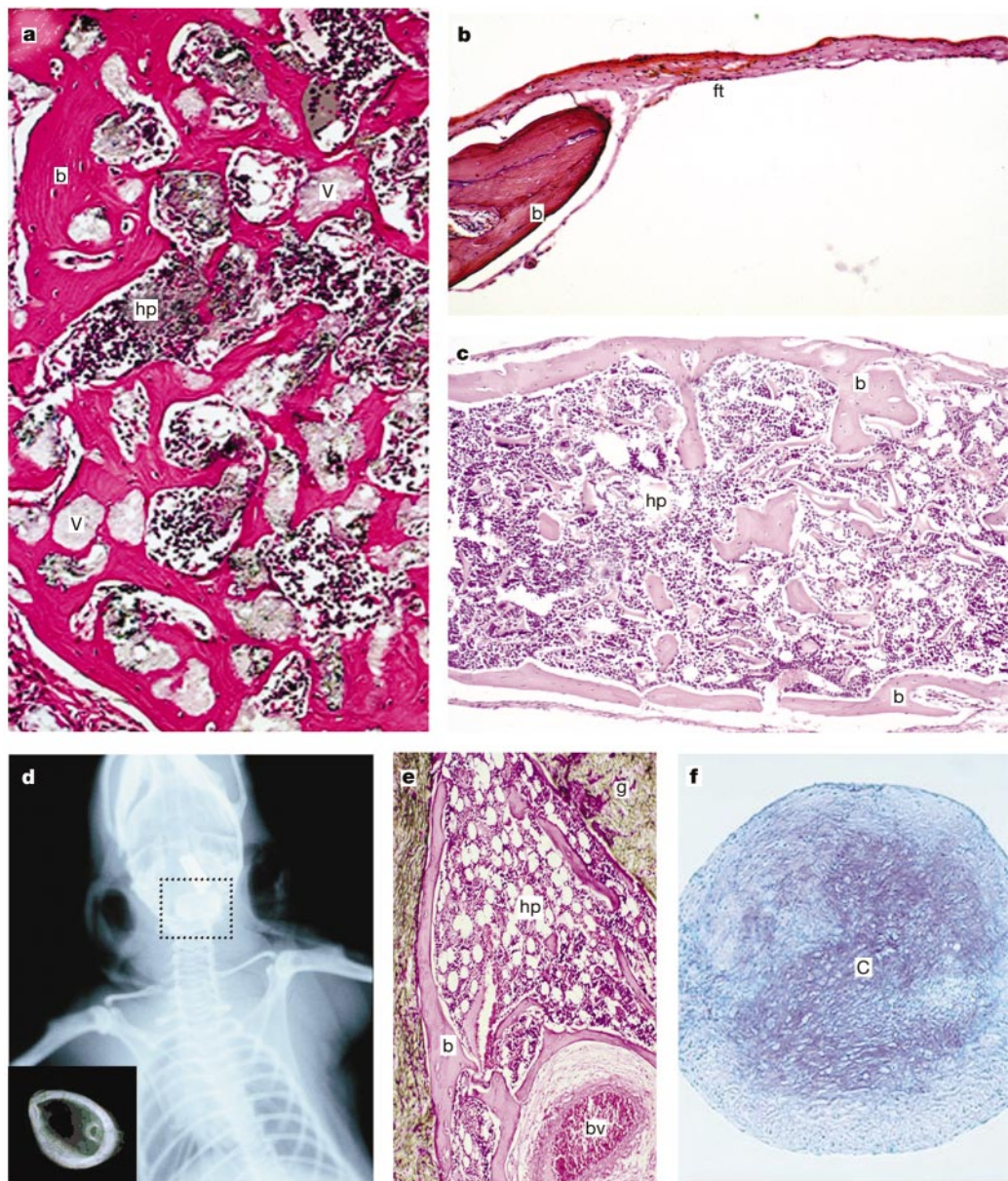
Figure 1 Regeneration of two-dimensional (skin) and three-dimensional (bone) tissues using stem cells. **a**, Skin autografts are produced by culturing keratinocytes (which may be sorted for p63, the recently described, epidermal stem cell marker) under appropriate conditions not only to generate an epidermal sheet, but also to maintain the stem cell population (holoclones). The epidermal sheet is then placed on top of a dermal substitute comprising devitalized dermis or bioengineered dermal substitutes seeded with dermal fibroblasts. Such two-dimensional composites, generated *ex vivo*, completely regenerate full-thickness wounds. **b**, Bone regeneration requires *ex vivo* expansion of marrow-derived skeletal stem cells and their attachment to three-dimensional scaffolds, such as particles of a hydroxyapatite/tricalcium phosphate ceramic. This composite can be transplanted into segmental defects and will subsequently regenerate an appropriate three-dimensional structure *in vivo*.

Table 1 Current approaches to tissue engineering

Stem cell-based tissue engineering		Non stem cell-based tissue engineering ¹	
Blood vessels ^{23,29}	Liver ³¹	Bladder	Meniscus
Bone ^{14–17*}	Pancreas ⁴¹	Cartilage (ear, nose and joints)*	Oral mucosa
Cartilage ³⁹	Nervous tissue ^{42*}	Heart valves	Salivary gland
Cornea ^{7,8*}	Skeletal muscle ^{24,27}	Intestine	Trachea
Dentin ⁴⁰	Skin ^{5,6*}	Kidney	Ureter
Heart muscle ^{28,29}			Urethra

*In clinical trials or clinical observational studies.

Figure 2 Bone regeneration by marrow-derived skeletal stem cells (SSCs). **a**, SSCs transplanted in conjunction with appropriate vehicles (v) such as hydroxyapatite/tricalcium phosphate generate an network of bone (b) and stroma composed of adipocytes and reticular cells that support haematopoiesis (hp). **b**, Critical size defects that do not heal spontaneously, such as those as in the calvaria, can be completely regenerated by SSC-vehicle constructs (**c**); ft, fibrous tissue. **d,e**, In addition, by wrapping SSC-vehicle constructs around an artery and vein (bv) in conjunction with Gore-Tex (g) to prevent collateral vascularization (dotted box in **d**), fully vascularized bone flaps of desired shape can be constructed. **f**, Under appropriate conditions, SSCs can also form cartilage (c) as demonstrated by alcian blue staining and type II collagen immunohistochemistry (not shown).



experimental evidence is available for limited engraftment of osteogenic progenitors that are infused (as is done with haematopoietic stem cells or HSCs) through the systemic circulation²⁰. But evidence for a biologically significant effect of the systemic infusion of SSCs is not available, and cure of systemic skeletal diseases using approaches borrowed simplistically from haematology is not in sight. Definitive preclinical work in animal models, compliant with strict criteria for assessment of engraftment, must be conducted before these types of procedures reach clinical trials¹¹, although preliminary clinical studies are underway²¹.

Another pertinent point is that renewal in a postnatal skeleton is markedly slower than in skin or blood (the whole skeleton being completely replaced only about three times in a lifetime, whereas the skin is replaced once a month). Consequently, we would expect replacement of skeletal tissue with infused SSCs to occur over longer timescales compared to rapidly self-renewing tissues, even if issues related to efficient cell delivery and systemic engraftment were resolved. Alternative means for local intervention in genetic diseases are conceivable. For example, engineered SSCs could be transplanted locally at the site of a clinical event (such as a fracture or deformity) in

genetic diseases of osteogenic cells, such as osteogenesis imperfecta (caused by mutations in genes encoding collagen type I) or fibrous dysplasia of bone (caused by activating missense mutations of the GNAS1 gene). But hopes for a systemic cure of genetic disorders of the skeleton will rely on more imaginative avenues. In some cases these might include *in utero* transplantation, whose feasibility with SSCs has been shown in preliminary studies²². As with the previously described techniques, caveats related to efficiency of engraftment and biological effects remain to be addressed.

Heterotopic stem cells for tissue engineering

The marrow is at centre stage for future technological developments in tissue engineering, not only as the only organ in which at least two types of stem cells (HSCs and SSCs) reside, but also as the organ in which progenitors for a number of distant tissues can be found. Recent studies indicate that the traditional wall separating the haematopoietic and mesodermal tissue systems and lineages is being demolished. Cells capable of regenerating blood vessels, skeletal muscle and cardiac muscle are found in the marrow^{23–25}. The unexpected potential for myogenesis and cardiomyogenesis has been

ascribed to both HSCs and stromal (mesenchymal) stem cells in the marrow^{26–29}. SSCs can perhaps give rise to neurons or glia³⁰, purified mouse HSCs can regenerate liver cells³¹, and cells able to regenerate bone are also found in blood³². Perhaps what we have referred to as the HSC is in itself much more — a true multipotent stem cell with transgerminal potentials^{2,31}, normally devoted to haematopoiesis as a result of local cues.

Marrow cells offer the advantage of being easily harvested and cultured from an adult organism, and the HSC can be isolated and purified *ex vivo*. Although most of these applications are a long way from any immediate clinical use, these studies raise basic scientific issues that are far from being settled or rationalized, similar, for example, to the issue of 'plasticity' of postnatal stem cells^{10,11}. They provide insight on how different the scene of tissue engineering could be in the relatively near future. Beyond theoretical considerations, and pending further experimental proof where needed, the existence of heterotopic and pleiotropic stem cells in the bone marrow has obvious practical implications for the future of stem cell therapy that should not be missed.

Reconstruction versus correction

Delivery of factors that would stimulate stem cells *in situ* to initiate a process leading to regeneration rather than scar formation has long been pursued. But success has been limited owing to problems of dosage, lack of full activity of recombinant factors, and inability to sustain a factor's presence for an appropriate length of time. To overcome these problems, 'gene-activated matrices' are being investigated that comprise plasmids coding for factors in a variety of delivery vehicles. These would transduce cells *in vivo* to bring about appropriate regeneration, and trials aimed at osteoporotic fractures are being planned³³. In these attempts, the traditional principles governing the use of carriers merge imaginatively with the frontier of genetic engineering. No doubt, if reconstruction was the initial frontier of tissue engineering, genetic correction is the next. As stem cells are the critical ingredient in tissue regeneration, they are also the critical targets of any strategy aimed at correcting a genetic defect.

Advances in our ability to genetically manipulate cells *ex vivo* using viral and non-viral transducing agents has circumvented many of the potential hazards of direct gene transfer in humans³⁴. Successful stable transduction of holoclone-generating epidermal cells with retroviral vectors has paved the way to the first trial of gene therapy for a devastating skin disease, junctional epidermolysis bullosa, a recessive disease caused by a malfunctioning laminin molecule³⁵. However, in dominant-negative diseases, a mutant gene must be silenced, and the techniques for doing this are in their infancy. Homologous recombination and site-directed mutagenesis of stem cells *ex vivo* would be the ultimate goal^{36,37}, but antisense RNA, RNA interference and hammerhead and hairpin RNAses offer intermediate strategies to at least prevent expression of a mutant protein³⁸.

With the prospect of stem cell-mediated gene therapy, the very definition of tissue engineering evolves into engineering of tissue function. Today, stem cell-based approaches to tissue reconstruction open unpredicted applicative (and market) opportunities. Although this discussion has been limited to systems where extensive preclinical and clinical studies have already been conducted, more exciting avenues are in sight in many areas. But enthusiasm over what unquestionably represents a markedly innovative technique with huge therapeutic potential must be balanced against stringent standards of scientific and clinical investigation. In addition, we must remain aware of the wide range of basic and applied issues associated with each system, with the targeted problems, and with the predicted solutions. □

1. Stock, U. A. & Vacanti, J. P. Tissue engineering: current state and prospects. *Annu. Rev. Med.* **52**, 443–451 (2001).
2. Lagasse, E., Shizuru, J. A., Uchida, N., Tsukamoto, A. & Weissman, I. L. Toward regenerative medicine. *Immunity* **14**, 425–436 (2001).

3. Barrandon, Y. & Green, H. Three clonal types of keratinocyte with different capacities for multiplication. *Proc. Natl Acad. Sci. USA* **84**, 2302–2306 (1987).
4. Pellegrini, G. *et al.* p63 identifies keratinocyte stem cells. *Proc. Natl Acad. Sci. USA* **98**, 3156–3161 (2001).
5. Ruzsac, Z. & Schwartz, R. A. Modern aspects of wound healing: an update. *Dermatol. Surg.* **26**, 219–229 (2000).
6. Pellegrini, G. *et al.* The control of epidermal stem cells (holoclones) in the treatment of massive full-thickness burns with autologous keratinocytes cultured on fibrin. *Transplantation* **68**, 868–879 (1999).
7. Pellegrini, G. *et al.* Long-term restoration of damaged corneal surfaces with autologous cultivated corneal epithelium. *Lancet* **349**, 990–993 (1997).
8. Tsai, R. J., Li, L. M. & Chen, J. K. Reconstruction of damaged corneas by transplantation of autologous limbal epithelial cells. *N. Engl. J. Med.* **343**, 86–93 (2000).
9. Friedenstein, A. J., Piatetzky, S. II & Petrakova, K. V. Osteogenesis in transplants of bone marrow cells. *J. Embryol. Exp. Morphol.* **16**, 381–390 (1966).
10. Bianco, P. & Gehron Robey, P. Marrow stromal stem cells. *J. Clin. Invest.* **105**, 1663–1668 (2000).
11. Bianco, P., Riniucci, M., Gronthos, S. & Robey, P. G. Bone marrow stromal stem cells: nature, biology, and potential applications. *Stem Cells* **19**, 180–192 (2001).
12. Langstaff, S. *et al.* Resorbable bioceramics based on stabilized calcium phosphates. Part I: rational design, sample preparation and material characterization. *Biomaterials* **20**, 1727–1741. (1999).
13. Hutmacher, D. W. Scaffolds in tissue engineering bone and cartilage. *Biomaterials* **21**, 2529–2543 (2000).
14. Bruder, S. P., Kraus, K. H., Goldberg, V. M. & Kadiyala, S. The effect of implants loaded with autologous mesenchymal stem cells on the healing of canine segmental bone defects. *J. Bone Joint Surg. Am.* **80**, 985–996 (1998).
15. Krebsbach, P. H., Mankani, M. H., Satomura, K., Kuznetsov, S. A. & Robey, P. G. Repair of craniofacial defects using bone marrow stromal cells. *Transplantation* **66**, 1272–1278 (1998).
16. Petite, H. *et al.* Tissue-engineered bone regeneration. *Nature Biotechnol.* **18**, 959–963 (2000).
17. Kon, E. *et al.* Autologous bone marrow stromal cells loaded onto porous hydroxyapatite ceramic accelerate bone repair in critical-size defects of sheep long bones. *J. Biomed. Mater. Res.* **49**, 328–337 (2000).
18. Quarto, R. *et al.* Repair of large bone defects with the use of autologous bone marrow stromal cells. *N. Engl. J. Med.* **344**, 385–386 (2001).
19. Mankani, M. H. *et al.* Pedicled bone flap formation using transplanted bone marrow stromal cells. *Arch. Surg.* **136**, 263–270 (2001).
20. Hou, Z. *et al.* Osteoblast-specific gene expression after transplantation of marrow cells: implications for skeletal gene therapy. *Proc. Natl Acad. Sci. USA* **96**, 7294–7299 (1999).
21. Horwitz, E. M. *et al.* Transplantation and therapeutic effects of bone marrow-derived mesenchymal cells in children with osteogenesis imperfecta. *Nature Med.* **5**, 309–313 (1999).
22. Liechty, K. W. *et al.* Human mesenchymal stem cells engraft and demonstrate site-specific differentiation after *in utero* transplantation in sheep. *Nature Med.* **6**, 1282–1286 (2000).
23. Kocher, A. A. *et al.* Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function. *Nature Med.* **7**, 430–436 (2001).
24. Ferrari, G. *et al.* Muscle regeneration by bone marrow-derived myogenic progenitors. *Science* **279**, 1528–1530 (1998).
25. Orlic, D. *et al.* Bone marrow cells regenerate infarcted myocardium. *Nature* **410**, 701–705 (2001).
26. Wakitani, S., Saito, T. & Caplan, A. I. Myogenic cells derived from rat bone marrow mesenchymal stem cells exposed to 5-azacytidine. *Muscle Nerve* **18**, 1417–1426 (1995).
27. Gussoni, E. *et al.* Dystrophin expression in the *mdx* mouse restored by stem cell transplantation. *Nature* **401**, 390–394 (1999).
28. Makino, S. *et al.* Cardiac myocytes can be generated from marrow stromal cells *in vitro*. *J. Clin. Invest.* **103**, 697–705 (1999).
29. Jackson, K. A. *et al.* Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *J. Clin. Invest.* **107**, 1395–1402 (2001).
30. Azizi, S. A., Stokes, D., Augelli, B. J., DiGirolamo, C. & Prockop, D. J. Engraftment and migration of human bone marrow stromal cells implanted in the brains of albino rats—similarities to astrocyte grafts. *Proc. Natl Acad. Sci. USA* **95**, 3908–3913 (1998).
31. Lagasse, E. *et al.* Purified hematopoietic stem cells can differentiate into hepatocytes *in vivo*. *Nature Med.* **6**, 1229–1234 (2000).
32. Kuznetsov, S. A. *et al.* Circulating skeletal stem cells. *J. Cell Biol.* **153**, 1133–1140 (2001).
33. Bonadio, J. Tissue engineering via local gene delivery: update and future prospects for enhancing the technology. *Adv. Drug Deliv. Rev.* **44**, 185–194 (2000).
34. Asahara, T., Kalka, C. & Isner, J. M. Stem cell therapy and gene transfer for regeneration. *Gene Ther.* **7**, 451–457 (2000).
35. Dellambra, E. *et al.* Toward epidermal stem cell-mediated *ex vivo* gene therapy of junctional epidermolysis bullosa. *Hum. Gene Ther.* **11**, 2283–2287 (2000).
36. Lanzov, V. A. Gene targeting for gene therapy: prospects. *Mol. Genet. Metab.* **68**, 276–282 (1999).
37. Li, J. & Baker, M. D. Mechanisms involved in targeted gene replacement in mammalian cells. *Genetics* **156**, 809–821 (2000).
38. Jen, K. Y. & Gewirtz, A. M. Suppression of gene expression by targeted disruption of messenger RNA: available options and current strategies. *Stem Cells* **18**, 307–319 (2000).
39. Johnstone, B. & Yoo, J. U. Autologous mesenchymal progenitor cells in articular cartilage repair. *Clin. Orthop.* **S156–S162** (1999).
40. Gronthos, S., Mankani, M., Brahimi, J., Robey, P. G. & Shi, S. Postnatal human dental pulp stem cells (DPSCs) *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **97**, 13625–13630 (2000).
41. Ramiya, V. K. *et al.* Reversal of insulin-dependent diabetes using islets generated *in vitro* from pancreatic stem cells. *Nature Med.* **6**, 278–282 (2000).
42. Bjorklund, A. Cell replacement strategies for neurodegenerative disorders. *Novartis Found. Symp.* **231**, 7–15 (2000).

Acknowledgements

The support of Telethon Fondazione Onlus Grant (to P.B.) is gratefully acknowledged.

Reprogramming of genome function through epigenetic inheritance

M. Azim Surani

Wellcome CRC Institute of Cancer and Developmental Biology, and Physiological Laboratory, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK (e-mail: as10021@mole.bio.cam.ac.uk)

Most cells contain the same set of genes and yet they are extremely diverse in appearance and functions. It is the selective expression and repression of genes that determines the specific properties of individual cells. Nevertheless, even when fully differentiated, any cell can potentially be reprogrammed back to totipotency, which in turn results in re-differentiation of the full repertoire of adult cells from a single original cell of any kind. Mechanisms that regulate this exceptional genomic plasticity and the state of totipotency are being unravelled, and will enhance our ability to manipulate stem cells for therapeutic purposes.

Development is a remarkably orderly process — it begins with a totipotent zygote and ends with an array of specific, differentiated cell types in adults. About 40,000 genes are needed to build a human being possessing ~200 histologically distinct cell types, and these categories can be subdivided further into a myriad of specialized cell

types. They fulfil precise functions that are as diverse as mounting a defence against diseases, regulating energy input–output and building neural networks, so allowing us to interact with our environment.

Once a cell is fully differentiated, this state is strikingly stable. Regardless of how different a neuron is from a hepatocyte, most cells retain an intact genome with the full complement of genes that are present at the beginning in the zygote. This simple concept of profound significance for development had its origin in the work of Spemann. The distinguishing features of cells arise from an orderly selection of genes that are expressed while the rest are switched off.

The genetic network that controls developmental decisions is beginning to be defined. The ability to acquire and inherit gene-expression patterns efficiently is also crucial to the individual history of cell differentiation. There are potential mechanisms that can allow differentiated cells to perpetuate the ‘molecular memory’ of the developmental decisions that created it. We know that this occurs without alterations or deletion of any DNA sequences, but rather by epigenetic mechanisms, which propagate appropriate patterns of gene expression (Fig. 1). These mechanisms involve heritable but potentially reversible modifications of DNA, primarily methylation of CpG (cytosine–guanine) dinucleotide¹. The binding of specific protein complexes to DNA also occurs to form stable and heritable chromatin structures that ensure efficient silencing of genes^{2,3} that are no longer required for determination of cell fate, allowing expression of only those genes that define properties of specific, differentiated cell types.

Although the mechanisms that perpetuate cell memory are naturally robust and reliable, they can be erased under some circumstances. The most pronounced manifestation of this erasure occurs when a differentiated somatic nucleus is transplanted back into an oocyte, which results in the restoration of totipotency^{4–7}. The reconstituted egg can then progress forward to generate a new organism that is a genetic copy or a clone of the individual nuclear donor. While not an efficient process, it is remarkable that it occurs at all; importantly, however, this establishes the principle that epigenetic states are reversible.

Mammalian genomes have an additional layer of epigenetic information referred to as genomic imprints, so called because they carry a molecular memory of their parental origin that is acquired in the germ line (Fig. 1). All our genomes therefore contain these distinct maternal and

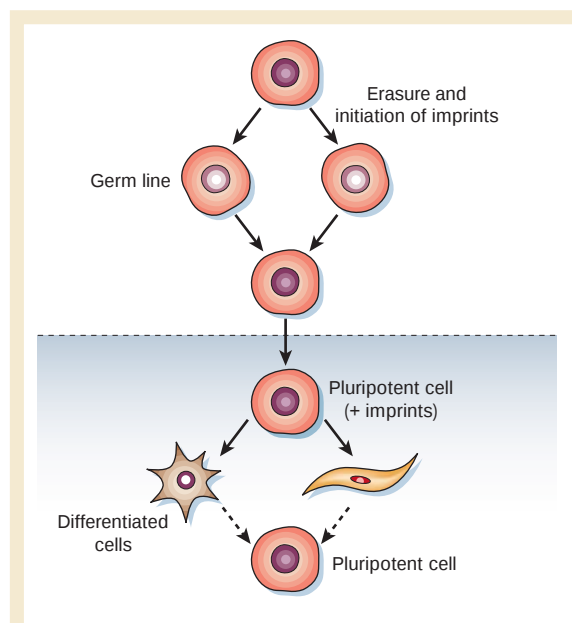


Figure 1 Epigenetic states are potentially reversible. Most cells contain the same set of genes, but their phenotype can vary according to which genes are expressed and repressed. Alterations in gene-expression patterns, without changes in DNA sequences, are referred to as epigenetic mechanisms. Epigenetic mechanisms make it possible to restore pluripotency to a differentiated cell, and a differentiated cell can also undergo transdifferentiation resulting in a pronounced change in its appearance and function. Mammalian genomes contain an additional layer of epigenetic information referred to as parental ‘imprints’. These imprints are erased and re-initiated normally in the germ line, and passed on to the offspring in which they survive into adulthood. Parental imprints also regulate gene expression and confer functional differences on parental genomes during development. Parental imprints can undergo changes without affecting the fundamental property of pluripotency.

paternal 'imprints' that are inherited after fertilization by embryos and endure thereafter into adulthood^{8,9}. These modifications, which are recognized as differential methylation of specific DNA sequences in sperm and oocytes, regulate expression of imprinted genes, which confer functional differences between parental genomes during development. Thus parental genomes exhibit an epigenetic asymmetry at fertilization, which persists throughout life. Furthermore, while the overall epigenetic state of the genome changes markedly during development and differentiation of cells, the parental imprints remain relatively stable.

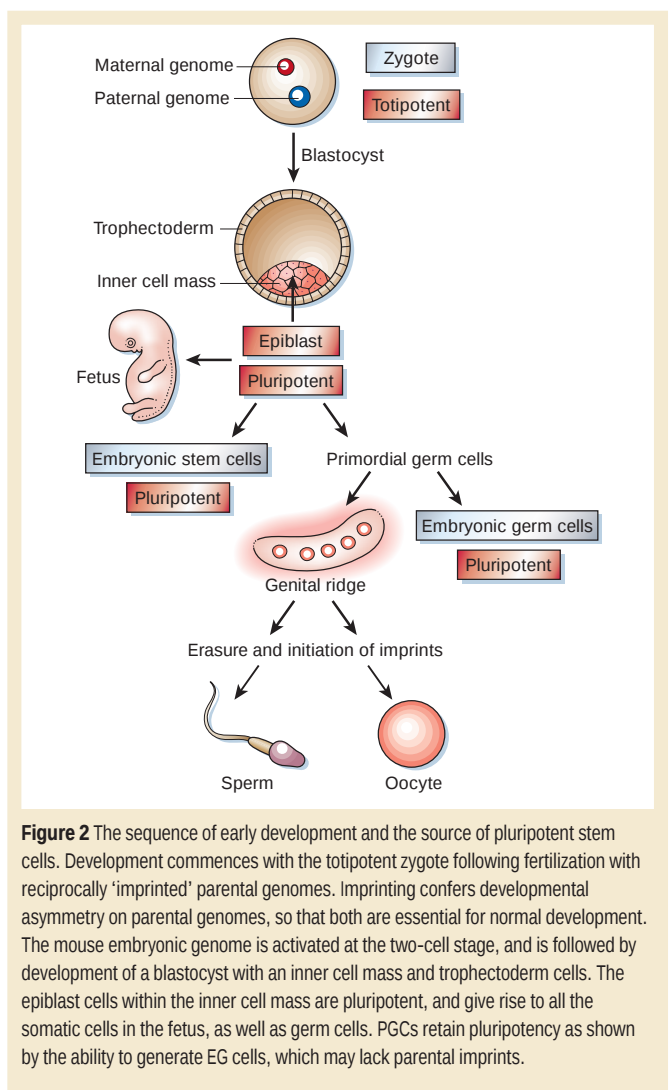
Reprogramming of genomes during imprinting in the germ line requires a stepwise cycle of erasure and re-initiation of imprints. A less well explored but a highly significant consequence of genomic imprinting is that the oocyte cytoplasmic factors have apparently evolved and acquired complex properties in mammals that are required to enhance and maintain the epigenetic asymmetry between parental genomes in the zygote (refs 8, 10–13, and K. Arney *et al.* unpublished data). These factors could have important consequences for reprogramming of a somatic nucleus to totipotency when transplanted into the oocyte. One key objective in this field is to gain a detailed knowledge of the mechanisms involved in the erasure of existing epigenetic states and establishment of new modifications for totipotency and during imprinting. These studies will allow us to assess more precisely events associated with reprogramming of somatic nuclei to a pluripotent or a totipotent state. The analysis of epigenetic mechanisms involved is also crucial for our ability to manipulate pluripotent stem cells and for the derivation of a range of differentiated cell types from pluripotent embryonic stem cells.

Epigenetic asymmetry between parental genomes

One of the main consequences of genomic imprinting and epigenetic asymmetry is that, whereas oocytes are potentially totipotent in many organisms, this is not so in mammals. This is because the maternal genome is epigenetically modified in the germ line to contain only the maternal 'imprints', which will normally result in the repression of certain maternally inherited imprinted genes. A paternal genome is essential to 'rescue' the oocyte, as the maternal genes are imprinted reciprocally to paternal imprints^{14,15}. So both parental genomes are needed for normal development — the paternal genome is relatively more important for development of the extraembryonic tissues, such as the trophoblast, whereas the maternal genome apparently has a greater influence on development of the embryo proper.

So far, about 45 imprinted genes have been identified in mice and humans^{8,9,16}. Imprinted genes may regulate some of the crucial aspects of mammalian physiology associated with reproduction, placenta, energy homeostasis, lactation and behaviour^{16–18}. For example, *Igf2*, which encodes a fetal insulin-like growth factor 2, is repressed in the maternal genome and active only in the paternal genome¹⁹. Other genes are repressed in the paternal genome and active in the maternal genome^{8,9,16} (for full details, see www.mgu.har.mrc.ac.uk). Some of the anomalies encountered in cloned embryos suggest disruption of imprinted gene expression²⁰.

Imprinted genes are often organized in clusters, sometimes in the megabase-range chromosomal regions containing key control elements — the differentially methylated regions (DMRs)^{8,9,16}. DMRs are CpG rich and subject to epigenetic modifications. These imprinting control regions are often complex with multiple functions acting to repress genes when methylated, or serving as boundary elements when unmethylated (the boundary element^{21,22} indirectly affects expression of neighbouring genes). Some DMRs also function as silencer elements when unmethylated²³, a function that is apparently abolished when the DMR is methylated. In other instances, a DMR is associated with the expression of an antisense transcript whose expression in turn ensures repression of the upstream gene²⁴. The result, in all cases, is to ensure monoallelic expression of imprinted genes. Such complex organization of imprinted genes means that any disruption of such clusters through chromosomal translocations or



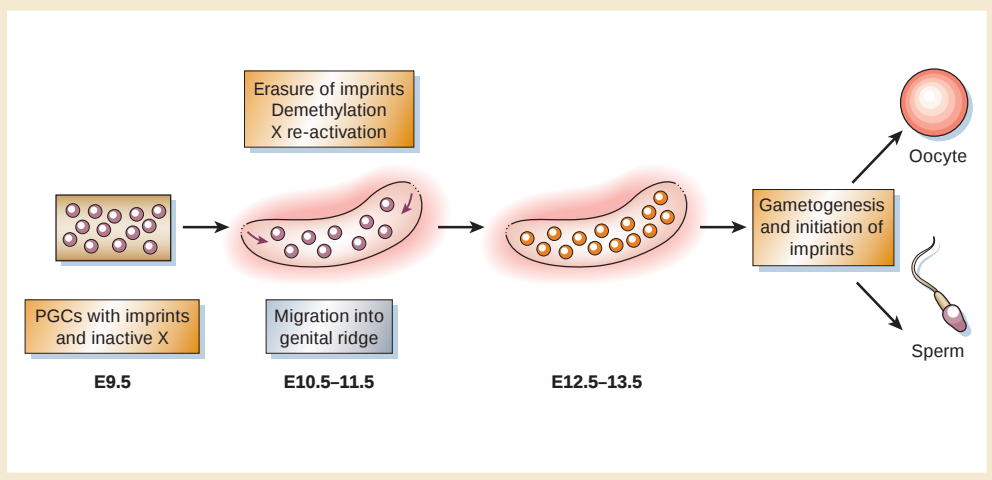
epigenetic mutations can cause anomalous expression of many genes within the cluster, and this accounts for some of the diseases associated with growth, neurogenetic disorders and diabetes^{8,9,16}.

Genomic imprinting probably accompanied mammalian evolution²⁵, and the evolution of placentation and viviparity in mammals resulted in significant changes in early development. One striking aspect is the emergence of trophoblast cells, which are essential for blastocyst implantation and as such are the first differentiated cell type to form during development. Also within the blastocyst are the pluripotent epiblast cells, the precursor of embryonic stem (ES) cells (Fig. 2). Gastrulation commences relatively late after implantation in response to signals that emanate from the trophoblast and primary endoderm cells. Another feature of mammalian development is that the oocytes are relatively small and lack the cytoplasmic determinants of development commonly encountered in other organisms, including those for the germ cell lineage. As a result, there is relatively early activation of the embryonic genome, which in the mouse occurs at the two-cell stage (Fig. 2).

The initiation of imprinting is confined to the germ line, first with the erasure of existing imprints in primordial germ cells (PGCs)^{26–29}, followed by the initiation of a new set of imprints in the male and female germ lines. It should be noted that whereas methylation of DMRs for some genes results in their repression, in other instances (for example, the *Igf2r* gene), methylation is essential for gene activation^{8,9,16}. In most instances, methylation of DMRs occurs predominantly in the female germ line. Nuclear transplantation studies between developing oocytes have shown that the maternal imprints

Figure 3 Reprogramming in germ cells.

Primordial germ cells in E9.5 embryos have the full complement of parental imprints. Upon the entry of PGCs into the genital ridge at E10.5–E11.5, extensive epigenetic modifications commence, perhaps in response to signalling molecules (purple arrows) from somatic cells. These epigenetic modifications include genome-wide demethylation, reactivation of the inactive X chromosome and erasure of imprints. By E13.5, definitive male and female gonads are formed and the PGCs are devoid of parental imprints. New sex-specific imprints are introduced later during gametogenesis, and are detected in mature sperm and oocytes.



are acquired at the time when oocytes resume growth from a quiescent state prior to ovulation^{30–32}. At this time, the X chromosome also acquires an imprint for the non-random inactivation of the paternal X chromosome in trophoblast and primary endoderm cells²⁹. The precise mechanism by which *de novo* methylation of DMRs occurs is not yet known, but it should involve some germline-specific factors acting in conjunction with DNA methyltransferase enzymes.

Genomic reprogramming in the germ line

Two important properties of PGCs are that they retain pluripotency (reviewed by Donovan and Gearhart, pages 92–97), and they are endowed with an exceptional capacity for epigenetic modifications of the genome. A unique feature is their ability to erase parental imprints, which shows that epigenetic modifications associated with imprinting can occur independently of the genomic status concerning pluripotency.

The pluripotent germ line

Elaborate transcriptional regulation is generally associated with the founding of germ cells to prevent them from acquiring a somatic cell fate. In mice, germ cells originate from the proximal epiblast cells of the embryonic day 6.5 (E6.5) egg cylinder. Cells migrate to the posterior proximal region where a founder population of ~45 PGCs is detected by E7.2 (refs 33, 34). It is the proximal location of the epiblast cells that is critical for germ cell fate, rather than any intrinsic properties of these cells³⁵. The specification of germ cell lineage depends on signals, such as bone morphogenetic protein (BMP)-4 and BMP8b (refs 36, 37), originating from the extraembryonic ectoderm in contact with the proximal epiblast. PGCs can be generated *in vitro* by combining epiblast with extraembryonic tissues³⁸ and, in principle, it should also be possible to generate PGCs from ES or embryonic germ (EG) cells. *Oct4*, a gene expressed in all totipotent and pluripotent cells in mammals^{39,40}, has an enhancer that is required for its expression predominantly in germ cells⁴⁰. There are, however, additional genes involved in the specification of the germ cell fate, and these are also likely to be crucial for pluripotency in general, including pluripotent stem cells (M. Saitou, S. C. Barton and M.A.S., unpublished data). Nevertheless, PGCs cannot participate in early development if re-introduced into blastocysts, unlike ES cells that can differentiate into a wide variety of somatic cells and germ cells. It may be that PGCs do not respond to signalling molecules, or that they are transcriptionally repressed. That PGCs are pluripotent is illustrated by the ability to derive EG cells from them^{41,42}, which share many common properties with ES cells. Precisely how a PGC reverts to a pluripotent EG cell remains unknown.

Epigenetic modifications in germ cells

When PGCs begin migration into the genital ridge at E9.5–10.5, they contain genomic imprints, and one of the two X chromosomes

is inactive in female PGCs (ref. 43, and P. Hajkova *et al.*, unpublished data). Pronounced epigenetic modifications commence with the entry of PGCs into the genital ridge (Fig. 3). There is rapid and possibly active genome-wide demethylation in both male and female PGCs, resulting in the erasure of imprints (refs 28, 30, 32, 44, and P. Hajkova *et al.*, unpublished data). This is accompanied by reactivation of the inactive X chromosome in female germ cells⁴³. The same mechanism also erases any aberrant epigenetic modifications, so preventing the inheritance of epimutations, which consequently occurs very rarely⁴⁵. The precise mechanism responsible for epigenetic erasure and demethylation in PGCs is as yet unclear.

Concerning the timing of epigenetic modifications in PGCs, there are at least two possibilities. First, these epigenetic modifications may be triggered in PGCs by a signal from somatic cells when they enter the genital ridge at E10.5–E11.5, which at this stage of development is undifferentiated and identical in both male and female embryos. Alternatively, the erasure of imprints might occur at a specific time and be regulated by a developmental clock. Whatever determines the timing of these events, PGCs by E13.5 possess an equivalent epigenetic state with erased imprints^{26,27,44}, and male and female gonads become distinguishable with distinct phenotypes. The erasure of imprints is also observed in EG cells²⁸. Here it occurs precociously, being found in EG cells derived from PGCs before their entry into the genital ridge, and could be due to culture of PGCs *in vitro*. It is important to note that ES cells do not show the same property for the erasure of imprints (see below). When the imprint-free EG cells are introduced into blastocysts to generate chimaeras, they can cause developmental anomalies, such as aberrant growth and skeletal abnormalities²⁸. The initiation of new imprints occurs subsequently during gametogenesis, and primarily during oogenesis.

Other forms of genomic modifications probably occur in PGCs, including restoration of telomeres to their optimum size and DNA repair generally⁴⁶. In single-cell organisms such as yeast, these functions reside within all cells, but in multicellular organisms, some of these functions occur either optimally or exclusively in the germ line. For example, the levels of telomerase are low in somatic cells, and this contributes to their ageing and senescence. Further work is needed to determine the capacity of the mammalian germ line to deal with DNA damage.

Genomic reprogramming in the zygote

Following imprinting in the germ line, the parental genomes exhibit epigenetic asymmetry at fertilization. These epigenetic differences are both maintained and enhanced in the zygote. During 0–5 hours post fertilization (h.p.f.), parental chromosomes can interact directly with maternally inherited cytoplasmic factors in the oocyte. Thereafter, the pronuclear membrane forms which can regulate access of

oocyte cytoplasmic factors to the parental genomes (Fig. 4). During the initial 0–5 h.p.f., the parental genomes exhibit dramatic epigenetic differences. The paternal genome undergoes marked demethylation while the maternal genome, which contains most of the methylation marks associated with imprints, undergoes further *de novo* methylation^{11,12,47}. Species where imprinting is unknown do not show such differential methylation of parental genomes during early development. Demethylation of the paternal genome may be essential to make it compatible for early activation of the embryonic genome. Alternatively, demethylation of the paternal genome may have accompanied evolution of developmental asymmetry between parental genomes¹³.

The oocyte cytoplasm has often been deployed to discriminate against the paternal genome during the evolution of diverse reproductive strategies and for generating interspecific barriers, which may have been the case during the evolution of genomic imprinting. Disruption of imprinting in interspecific mammalian hybrids of the deer mouse, *Peromyscus maniculatus*, may be due to nuclear–cytoplasmic incompatibility⁴⁸. Furthermore, certain oocyte cytoplasmic modifiers can induce epigenetic modifications of target loci, subsequently rendering them inactive by DNA methylation⁴⁹. Such incompatibility may also arise with transplantation of the somatic nucleus into oocytes. The response of a somatic nucleus is likely to be dictated both by its original state, and by how it reacts to the complex oocyte cytoplasmic factors.

There are a number of maternally inherited oocyte cytoplasmic factors with the potential to modify the epigenetic states of parental genomes and of the transplanted somatic nuclei. One such factor is the heterochromatin protein HP1, which may interact differentially with parental genomes, and with somatic nuclei depending on their existing epigenetic state (ref. 10, and K. Arney, unpublished data). HP1 can bind methylated histone H3 (meH3) via a chromodomain, and there is growing evidence indicating that this interaction can lead to *de novo* DNA methylation. Such interactions may account for the differential methylation of parental genomes in the zygote (K. Arney *et al.*, unpublished data). HP1–meH3 interaction also provides a mechanism for the inheritance of the newly established epigenetic states, a mechanism that is apparently widely conserved in many organisms, including yeast^{2,3}. The histone methyltransferase activity is intrinsic to the SET domain of *Su(var)3-9*, a *Polycomb Group*

(Pc-G) gene³. This mechanism might also be involved in the initiation of parental imprints during oocyte development.

The oocyte cytoplasm contains several other Pc-G proteins, with the products of *Ezh2* and *eed* seeming particularly crucial for early mammalian development^{10,50}. EZH2 and EED, together with YY1 (mammalian homologue of *Drosophila pleiohmiotic*), form a complex with histone deacetylases (HDACs) 1 and 2 that has the potential to mediate transcriptional repression⁵¹. It is significant that loss of function of any of the three Pc-G genes results in very early embryonic lethality^{50,52,53}. It is particularly striking that EZH2, EED and YY1 are present in the oocyte cytoplasm and translocate into pronuclei after fertilization (ref. 10, and S. Erhardt and K. Arney, unpublished data). EZH2, which also has the conserved SET domain⁵³, may potentially be involved in histone methylation. The role of meH3–HP1 interactions in inducing *de novo* methylation and as a heritable epigenetic mechanism has the potential to be important in pluripotent stem cells and their subsequent differentiation. For example, the loss of function of *Ezh2* is incompatible with the derivation of ES cells⁵³, which is particularly remarkable since ES cells lacking DNA methyltransferases, such as DNMT1, can survive and proliferate extensively⁵⁴.

DNMT1, which is also present in the oocyte, is known to be excluded from entry into pronuclei, which is why demethylation of the embryonic genome continues throughout pre-implantation development until the blastocyst stage^{26,47,55}. Genome-wide *de novo* methylation occurs later in early post-implantation embryos. Despite these marked changes in genomic DNA methylation, the parental imprints are preserved, possibly because of the specific properties of the DMRs. The transient entry of DNMT1 into the nuclei at the eight-cell stage seems to help to maintain the imprints⁵⁵.

Some aspects of transcriptional regulation may be shared between early embryos and PGCs. It is likely, for example, that *Ezh2* has a significant role in transcriptional regulation in the germ line. There are only two Pc-G genes known in *Caenorhabditis elegans*; both are homologues of *Ezh* and *eed*, and are critical for transcriptional repression. Loss of function of either of the two genes results in the loss of germ cell lineage in worms. Genome-wide demethylation is also observed in PGCs and in the paternal genome in the zygote^{11,12,56}, although it is important to note that whereas the imprints are erased in PGCs, this is not the case during embryonic development.

Reprogramming somatic nuclei in the oocyte

Transplantation of a somatic nucleus into the oocyte has the potential to restore totipotency to the somatic nucleus⁵⁷. This transformation probably involves erasure of the existing epigenetic state and a reversion to an embryonic pattern of gene expression. Both the erasure of DNA methylation and the initiation of new epigenetic modifications seem possible, based on studies on the zygote. The imprints have the potential to survive during reprogramming of somatic nuclei to totipotency^{11,12,47,58}, but are occasionally erased as well⁵⁷.

The interactions between oocyte cytoplasmic factors, such as HP1, and a somatic nucleus will depend in part on the epigenetic state of the donor nucleus when transplanted into oocytes. A donor nucleus is usually transplanted into a non-activated oocyte, which would initially involve direct interactions between the donor chromosomes and the oocyte cytoplasmic factors. Such interactions may result in both *de novo* methylation and demethylation of specific loci of donor nuclei, which may occur stochastically, resulting subsequently in aberrant patterns of gene expression and failure of development. In one recent study, the methylation status of the donor nucleus at the blastocyst stage was found to be either unchanged or aberrant compared to the control⁵⁹. Reprogramming of the X chromosome can occur, however, as the inactive X chromosome can be reactivated in the somatic nucleus, although the molecular memory of the inactive X chromosome is retained, resulting in its preferential inactivation in the trophectoderm⁶⁰. Although imprinted genes may be expected to remain largely unaffected during

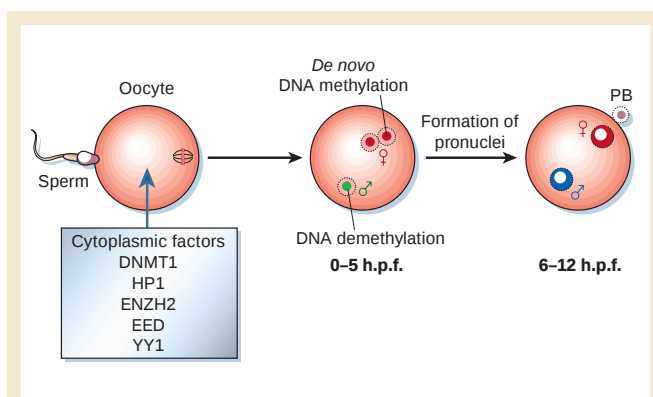


Figure 4 Reprogramming in the zygote. Cytoplasmic factors stored in the oocyte commence interactions with parental genomes after fertilization. Between 0 and 5 hours post fertilization (h.p.f.), the parental genomes display marked differences in epigenetic modifications, with the paternal genome undergoing demethylation and the maternal genome showing *de novo* methylation. This accentuates the epigenetic asymmetry between parental genomes. After the formation of pronuclei at approximately 6 h.p.f., further epigenetic modifications are regulated by the entry of cytoplasmic factors into the nuclei. For example, DNMT1 is excluded from entry into the nuclei. These cytoplasmic factors and others have the potential to modify the epigenetic state of somatic nuclei transplanted into the oocyte. PB, polar body.

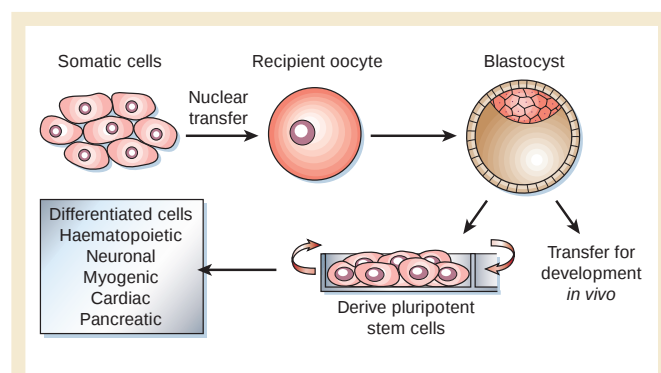


Figure 5 Reprogramming a somatic nucleus. When transplanted into an oocyte, a somatic nucleus may respond to the cytoplasmic factors and be reprogrammed back to totipotency. These cytoplasmic factors must be capable of erasing the 'molecular memory' that gives somatic cells their characteristic properties. It would also be necessary for the reprogrammed nucleus to switch off specific genes that are expressed by the somatic nucleus and initiate embryo-specific genes at the two-cell stage in the mouse. The reprogrammed genome generates pluripotent epiblast cells, and undergoes rapid transdifferentiation to generate trophectoderm cells. The extent of reprogramming can be judged by the development of the conceptus *in vivo*, as well as by the derivation of ES cells from blastocysts. These ES cells can potentially be induced to differentiate to generate the entire repertoire of adult cell types and germ cells.

reprogramming of somatic nuclei to totipotency, there are nevertheless some instances where imprints may be erased, which leads to fetal and placental growth anomalies⁵⁷. Many other phenotypic anomalies have been noted after live birth (for example, respiratory problems^{20,61}), although it is unclear if this is due to genetic or epigenetic anomalies.

Several additional factors influence survival after birth⁶¹, including the genetic background of the donor nucleus. A variety of somatic cell types of different ages have been examined for cloning efficiency, but this still remains at less than 3% (refs 62, 63). Clearly, the oocyte cytoplasmic factors are designed primarily to modify the distinct epigenetic states of parental genomes. A better understanding of both the mechanism of these epigenetic changes and the selection and prior modifications of donor nuclei might improve the outcome.

The oocyte-derived components must also contain other modifiers of chromatin. In amphibians, at least one chromatin-remodelling factor, the ATPase ISWI (a member of the SW12/SNF2 superfamily), erases TATA-binding protein from association with the nuclear matrix of somatic nuclei⁶⁴. This and other kinds of chromatin-remodelling activities must occur in mice as well. There are also factors, such as OCT4, that could have an essential role in restoring totipotency, whereas other factors might be required for early development until at least the activation of the embryonic genome, and possibly during pre-implantation development.

Another measure of reprogramming of somatic nuclei is the potential to restore the telomere size and to overcome senescence^{65–68}. Although there is some evidence that telomere length is restored in somatic nuclei after transplantation into oocytes, most DNA mutations in the somatic nucleus cannot be repaired. Although such mutations may be tolerated in a differentiated cell, they can be lethal at any stage of early development. This might also contribute to the low efficiency of development following transplantation of somatic nuclei into oocytes. It is significant that transplantation of nuclei from pluripotent ES cells into oocytes results in higher rates of development to term⁶⁹. Development of procedures to select and modify the epigenetic status of the donor nucleus might improve the frequency of normal development.

Whereas most attention has focused on how a somatic nucleus may acquire a totipotent state, it has been largely overlooked that the

nucleus must also be reprogrammed to undergo rapid transdifferentiation to generate the highly specialized trophectoderm cells by the fourth cleavage division in the mouse. This remarkable example of genomic plasticity is crucial for the subsequent embryonic development. The extraembryonic tissues in mammals are critical as the source of signalling molecules and must function optimally for differentiation of both embryonic somatic cells and for the establishment of the germ line.

Are imprints essential for development?

Because imprints and DNA methylation of the somatic nucleus may affect development, is it possible to obtain normal development if the asymmetry between parental genomes is removed by complete erasure of imprints? Evidence suggests that an imprint-free (and undermethylated) PGC nucleus transplanted to an oocyte is incapable of development to term, even though PGCs are pluripotent³⁰. A larger proportion of these reconstituted zygotes do develop to the blastocyst stage, but the resulting conceptuses fail to progress. In part, this results from an aberrant placental phenotype, which to some extent can be attributed to the loss of function of *Mash2*, a maternally expressed imprinted gene. Loss of function of other key genes such as *Igf2* occurs with the loss of imprints³⁰. There is clear evidence to show that the oocyte cannot initiate imprints if they are erased; imprints can only be initiated in the germ line³⁰.

The ability to derive ES cells from blastocysts after nuclear transfer can be used as a measure of nuclear reprogramming (Fig. 5). Individual blastocysts from various somatic cells have been used to derive ES cells, but the success rate is about 4%, compared to an average of 40–50% from normal blastocysts of the 129/Sv strain⁷⁰. Derivation of ES cells is apparently unaffected by the lack of imprints, as these cells were obtained from blastocysts generated with the imprint-free (and substantially demethylated) PGC nuclei³⁰. Remarkably, the frequency of ES cell derivation in this case was close to 100%, perhaps because PGCs are themselves pluripotent and substantially unmethylated (Y. Kato and M.A.S., unpublished data). It seems that the ability to restore overall pluripotency to somatic nuclei is feasible with aberrant or complete lack of imprints. Additionally, many specific mutations may have no effect on reprogramming of somatic nuclei to pluripotency and the subsequent derivation of ES cells, if they do not affect early development. The consequences of some genetic and epigenetic mutations may become evident only later when these cells are allowed to undergo differentiation towards specific cell types.

Reprogramming factors in pluripotent stem cells

The distinguishing feature of pluripotent ES or EG cells is that they exist in a transcriptionally permissive state⁷¹. Potentially, this allows indefinite self-renewal without the imposition of restrictions, at least until they are permitted to differentiate into diverse cell types. Then, they undergo progressive restrictions during differentiation, as occurs during development from the totipotent zygote.

Pluripotent stem cells themselves must possess factors with the ability to reprogramme somatic nuclei to pluripotency. As mammalian oocytes are immensely complex, and small in size and numbers, ES and EG cells may provide an alternative system to identify the critical factors and mechanisms necessary for pluripotency using a variety of biochemical, genetic and cell biological approaches. Fusion between pluripotent and somatic cells is the first step towards elucidating the potential of EG and ES cells to modify a somatic nucleus (Fig. 6). This approach, involving fusion between different cell types, has been used previously to examine aspects of gene regulation and genomic plasticity.

Reactivation of the inactive X chromosome occurs when thymocytes are fused with pluripotent embryonal carcinoma cells⁷². Both ES and EG cells apparently show dominant activities concerning reprogramming of somatic nuclei after cell fusion, presumably in response to *trans*-acting factors from the pluripotent cells. For example, fusion of a somatic nucleus to an ES cell results in reactivation of

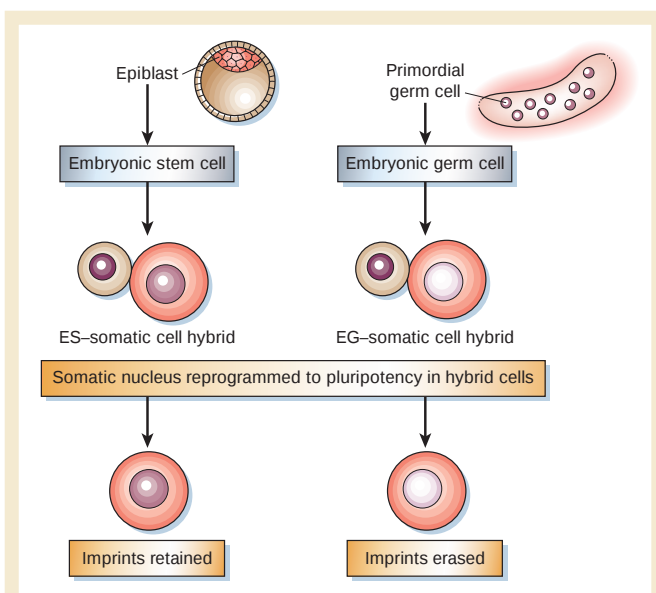


Figure 6 Reprogramming in ES and EG cells. A somatic nucleus when fused with an ES or EG cell undergoes extensive reprogramming to pluripotency. The *Oct4* gene, which is silent and methylated in somatic cells, is also reactivated as it undergoes demethylation. However, fusion with EG cells, but not ES cells, results in genome-wide demethylation and erasure of imprints from somatic nuclei, a property that is inherited from the precursor primordial germ cells. Pluripotent cells therefore contain factors for extensive modifications of somatic cells, which can confer pluripotency on somatic cells.

Oct4 in the somatic nucleus⁷³ (P. Western and M.A.S., unpublished data). Hence, when thymocytes carrying the *Oct4-GFP* transgene are fused with ES cells, the reporter expression was observed in hybrid cells within 24–48 h (possibly after 1–2 cell divisions), illustrating that transcriptional reactivation of *Oct4* occurs rapidly. As the *Oct4* gene is methylated in somatic cells, its expression in ES–somatic cell hybrids is accompanied by demethylation of the promoter region⁷³. Expression of the *Oct4* gene is confined to totipotent and pluripotent cells^{40,74}, and the expression of *Oct4* from the somatic nucleus is a clear reflection of reprogramming of this somatic nucleus. A similar response is observed when other types of somatic cells are fused with ES cells.

The dominant activity of an EG cell fused with a somatic cell results in the induction of similar epigenetic changes in the somatic cell, together with the restoration of the pluripotent state⁷⁵. However, EG cells also have the ability to erase parental imprints, a property they inherit from their germline precursor cells. This property is absent in ES cells, oocytes and early embryos. Thus, when fused with a thymocyte, EG cells induce erasure of parental imprints from the somatic nucleus. Furthermore, a silent imprinted allele of the *Mest* gene is reactivated after demethylation⁷⁵. Other changes to the somatic nucleus include repression of the thymocyte-specific transcript *Thy-1.2* and reactivation of the inactive X chromosome. The hybrid cells also exhibit pluripotency when introduced into blastocysts^{73,75}. The ability of ES and EG cells to reprogramme a somatic nucleus therefore reflects the properties inherited from their precursor cells, the epiblast and germ cells, respectively.

Perspective

Germ cells, stem cells and early embryos all exhibit pluripotency, but each cell type also displays certain unique properties. It is essential to identify the precise nature of the pluripotent state shared by these different cell types, including how it is attained and propagated and which genes confer pluripotency. At the same time, it is imperative to elucidate the mechanisms by which the epigenetic states are changed,

so resulting in profound effects on cell potency. Alterations to epigenetic modifications allow a switch in patterns of gene expression, which are central to genomic plasticity and transdifferentiation. While emphasizing the intrinsic nature of the switch, environmental factors also play a fundamental role *in vivo*, and the impact of these factors on epigenetic modifications must be determined. Many of the advances in understanding the intrinsic mechanisms by which embryonic cell fate is determined during development will prove informative for manipulating cell fates. The immense potential of stem cells will rely on understanding many of the critical mechanisms that regulate the appropriate selection of genes available in all cells. □

- Bird, A. P. & Wolffe, A. P. Methylation-induced repression—belts, braces and chromatin. *Cell* **99**, 451–454 (1999).
- Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromatin domain. *Nature* **410**, 120–124 (2001).
- Jenuwein, T. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266–273 (2001).
- Gurdon, J. The developmental capacity of nuclei taken from intestinal epithelial cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* **10**, 622–640 (1962).
- Wilmot, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
- Wakayama, T., Zuccotti, M., Johnson, K. R., Perry, A. C. F. & Yanagimachi, R. Full term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
- Kato, Y. *et al.* Eight calves cloned from somatic cells of a single adult. *Science* **282**, 2095–2098 (1998).
- Ferguson-Smith, A. C. & Surani, M. A. Imprinting and the epigenetic asymmetry between parental genomes. *Science* **293**, 1086–1089 (2001).
- Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature Rev. Genet.* **2**, 21–32 (2001).
- Arney, K. L., Erhardt, S., Drewell, R. A. & Surani, M. A. Epigenetic reprogramming of the genome—from the germ line to the embryo and back again. *Int. J. Dev. Biol.* **45**, 533–540 (2001).
- Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
- Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
- Reik, W. & Walter, J. Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. *Nature Genet.* **27**, 255–256 (2001).
- Surani, M. A. H., Barton, S. C. & Norris, M. L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548–550 (1984).
- McGrath, J. & Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179–183 (1984).
- Tilghman, S. M. The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell* **96**, 185–193 (1999).
- Lefebvre, L. *et al.* Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest*. *Nature Genet.* **20**, 163–169 (1998).
- Li, L. *et al.* Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. *Science* **284**, 330–333 (1999).
- DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849–859 (1991).
- Tamashiro, K. L. K., Wakayama, T., Blanchard, R. J., Blanchard, D. C. & Yanagimachi, R. Postnatal growth and behavioural development of mice cloned from adult cumulus cells. *Biol. Reprod.* **63**, 328–334 (2000).
- Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
- Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
- Constancia, M. *et al.* Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nature Genet.* **26**, 203–206 (2000).
- Lyle, R. *et al.* The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. *Nature Genet.* **25**, 19–21 (2000).
- Killian, J. K. *et al.* M6P/IGF2R imprinting evolution in mammals. *Mol. Cell* **5**, 707–716 (2000).
- Brandeis, M. *et al.* The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J.* **12**, 3669–3677 (1993).
- Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.* **6**, 705–714 (1992).
- Tada, T. *et al.* Epigenotype switching of imprintable loci in embryonic germ cells. *Dev. Genes Evol.* **207**, 551–561 (1998).
- Tada, T. *et al.* Imprint switching for non-random X-chromosome inactivation during mouse oocyte growth. *Development* **127**, 3101–3105 (2000).
- Kato, Y. *et al.* Developmental potential of mouse primordial germ cells. *Development* **126**, 1823–1832 (1999).
- Kono, T., Obata, Y., Yoshimizu, T., Nakahara, T. & Carroll, J. Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. *Nature Genet.* **13**, 91–94 (1996).
- Obata, Y. *et al.* Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesis. *Development* **125**, 1553–1560 (1998).
- Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found. Symp.* **182**, 68–84 (1994).
- McLaren, A. Signaling for germ cells. *Genes Dev.* **13**, 373–376 (1999).
- Tam, P. P. & Zhou, S. X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev. Biol.* **178**, 124–132 (1996).

36. Lawson, K. A. *et al.* Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).
37. Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol. Endocrinol.* **14**, 1053–1063 (2000).
38. Yoshimizu, T., Obinata, M. & Matsui, Y. Stage-specific tissue and cell interactions play key roles in mouse germ cell specification. *Development* **128**, 481–490 (2001).
39. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct-4. *Cell* **95**, 379–391 (1998).
40. Pesce, M., Gross, M. K. & Scholer, H. R. In line with our ancestors: Oct-4 and the mammalian germ. *Bioessays* **20**, 722–732 (1998).
41. Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847 (1992).
42. Resnick, J. L., Bixler, L. S., Cheng, L. & Donovan, P. Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550–551 (1992).
43. Tam, P. L., Zhou, S. X. & Tan, S.-S. X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development* **120**, 2925–2932 (1994).
44. Mann, J. R. Imprinting in the germ line. *Stem Cells* **19**, 289–294 (2001).
45. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23**, 314–318 (1999).
46. Meeker, A. K. & Coffey, D. S. Telomerase: a promising marker of biological immortality of germ, stem, and cancer cells. A review. *Biochemistry* **62**, 1323–1331 (1997).
47. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382 (1987).
48. Vrana, P. B., Guan, X. J., Ingram, R. S. & Tilghman, S. M. Genomic imprinting is disrupted in interspecific *Peromyscus* hybrids. *Nature Genet.* **20**, 362–365 (1998).
49. Allen, N. D., Norris, M. L. & Surani, M. A. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* **61**, 853–861 (1990).
50. Schumacher, A., Faust, C. & Magnuson, T. Positional cloning of a global regulator of anterior–posterior patterning in mice. *Nature* **384**, 648 (1996).
51. Satijn, D. P., Hamer, K. M., den Blaauwen, J. & Otte, A. P. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in *Xenopus* embryos. *Mol. Cell. Biol.* **21**, 1360–1369 (2001).
52. Donohoe, M. E. *et al.* Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell. Biol.* **19**, 7237–7244 (1999).
53. O'Carroll, D. *et al.* The Polycomb group gene *Ezh2* is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
54. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
55. Howell, C. Y. *et al.* Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **104**, 829–838 (2001).
56. Monk, M. & Harper, M. I. Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature* **281**, 311–313 (1979).
57. Solter, D. Mammalian cloning: advances and limitations. *Nature Rev. Genet.* **1**, 199–207 (2000).
58. Olek, A. & Walter, J. The pre-implantation ontogeny of the H19 methylation imprint. *Nature Genet.* **17**, 275–276 (1997).
59. Kang, Y.-K. *et al.* Aberrant methylation of donor genome in cloned bovine embryos. *Nature Genet.* **28**, 173–177 (2001).
60. Eggan, K. *et al.* X-Chromosome inactivation in cloned mouse embryos. *Science* **290**, 1578–1581 (2000).
61. Eggan, K. *et al.* Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl Acad. Sci. USA* **98**, 6209–6214 (2001).
62. Wakayama, T. & Yanagimachi, R. Mouse cloning with nucleus donor cells of different age and type. *Mol. Reprod. Dev.* **58**, 376–383 (2001).
63. Wakayama, T. & Yanagimachi, R. Effect of cytokinesis inhibitors, DMSO and the timing of oocyte activation on mouse cloning using cumulus cell nuclei. *Reproduction* **122**, 49–60 (2001).
64. Kikyo, N., Wade, P. A., Guschin, D., Ge, H. & Wolffe, A. P. Active remodeling of somatic nuclei in egg cytoplasm by the nucleosomal ATPase ISWI. *Science* **289**, 2360–2362 (2000).
65. Betts, D. *et al.* Reprogramming of telomerase activity and rebuilding of telomere length in cloned cattle. *Proc. Natl Acad. Sci. USA* **98**, 1077–1082 (2001).
66. Lanza, R. P. *et al.* Extension of cell life-span and telomere length in animals cloned from senescent somatic cells. *Science* **288**, 665–669 (2000).
67. Shiels, P. G. *et al.* Analysis of telomere lengths in cloned sheep. *Nature* **399**, 316–317 (1999).
68. Wakayama, T. *et al.* Cloning of mice to six generations. *Nature* **407**, 318–319 (2000).
69. Rideout, W. M. *et al.* Generation of mice from wild-type and targeted ES cells by nuclear cloning. *Nature Genet.* **24**, 109–110 (2000).
70. Wakayama, T. *et al.* Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. *Science* **292**, 740–743 (2001).
71. Smith, A. G. Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol.* (in the press).
72. Takagi, N., Yoshida, M. A., Sugawara, O. & Sasaki, M. Reversal of X-inactivation in female mouse somatic cells hybridized with murine teratocarcinoma stem cells in vitro. *Cell* **34**, 1053–1062 (1983).
73. Tada, M. Nuclear reprogramming of somatic cells by in vitro hybridisation with ES cells. *Curr. Biol.* (in the press).
74. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
75. Tada, M., Tada, T., Lefebvre, L., Barton, S. C. & Surani, M. A. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J.* **16**, 6510–6520 (1997).

Acknowledgements

I thank members of my laboratory for their valuable contributions and critical comments during the course of our studies, Barry Keverne for critical comments on the manuscript and the Wellcome Trust for generously supporting my work. Space constraints has prevented citation of a large number of key papers.

Ethical and social considerations of stem cell research

Anne McLaren

Wellcome/CRC Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK (e-mail: a.mclaren@welc.cam.ac.uk; am10021@hermes.cam.ac.uk)

Much recent interest has focused on whether stem cell therapy could alleviate or even cure common degenerative diseases. This has been accompanied by debate on the ethics of destructive research on early human embryos. Stem cells derived from various sources raise different ethical issues, but their contribution to medical research could be immense. Any use of stem cells should be subject to appropriate scientific and ethical review.

Every year millions of people suffer and eventually die from serious and largely incurable degenerative diseases of the nervous system (for example, Parkinson's disease, multiple sclerosis and stroke), heart (myocardial infarction), liver (hepatitis), pancreas (diabetes) and other organs. Stem cell therapy could alleviate or even cure some of these diseases — but what about the ethics, if the stem cells come from embryos? Any new, experimental medical treatment raises ethical issues for both doctors and patients, but research on embryonic stem (ES) cells raises in addition the ethical conflict between destructive human embryo research and the magnitude of the possible benefits. The huge scale of the clinical problem means that resource considerations cannot be ignored, given that equity is itself an ethical issue. These conflicts, and their social and legal implications, have stimulated discussions worldwide.

Stem cell derivation

Stem cells can be harvested not only from embryos, but also from adults (living or dead) or from fetuses. From the ethical point of view, the use of adult stem cells (Fig. 1) for therapy raises no new issues. Bone-marrow transfusion and organ transplantation are widely practised and regulated by guidelines. The use of fetal tissue, including fetal stem cells, is also regulated by guidelines, although people who regard elective termination of pregnancy as ethically unacceptable include some who would not countenance the use of tissue from an aborted fetus for research or treatment. A fetus cannot give informed consent. In addition, although adult and fetal stem cells may be multipotent¹ — that is, capable of generating many different cell types — they are present only in small numbers, so would not provide a realistic approach to cell therapy of common degenerative diseases unless they could be greatly amplified in culture while retaining a stable karyotype. Stem cells in the cord blood of one baby are sufficient to treat only a single patient.

Pluripotent embryonic stem cells derived from mice can not only be amplified indefinitely in culture, but they can also give rise to all the cell types found in the adult body, as reviewed by Donovan and Gearhart, pages 92–97. On their own, however, they cannot give rise to an embryo. Human stem cell lines thought to be pluripotent have been derived both from primordial germ cells developing in fetal tissue (embryonic germ (EG) cell lines), and from

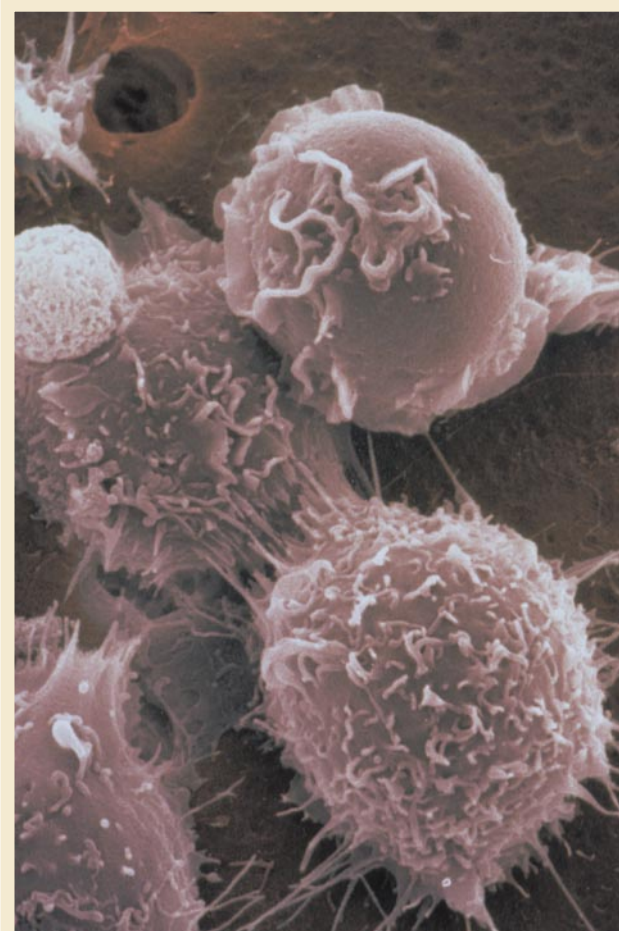


Figure 1 Multipotent bone-marrow stem cells form the precursors to various types of blood cell. (Image courtesy of A. Leonard/SPL.)

early pre-implantation embryos no longer required for infertility treatment (ES cell lines). Less is known about the potential of EG cells, and concerns have been expressed that the altered methylation status of the germ cell lineage might affect their therapeutic value. Their use would be regulated by the same guidelines as other fetal tissue. Informed consent is of course an ethical requirement for any donated tissue, but for potentially immortal stem cell lines, the information needs to take account both of

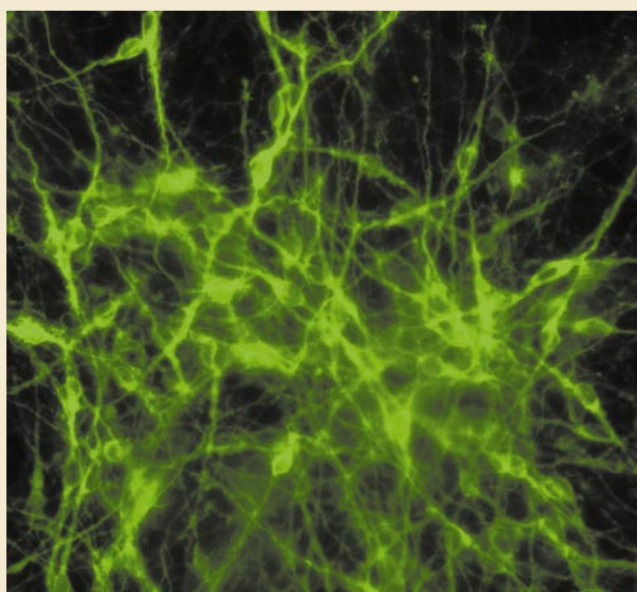


Figure 2 Tyrosine hydroxylase expression in dopaminergic neurons derived from mouse ES cells. (Image courtesy of J. Kim and R. McKay.)

the genetic implications and any arrangements that are being made concerning anonymity, and also of the fact that the cells may be used for treatment.

ES cell lines have been studied for many years in mice, where they can be induced to develop into, for example, dopaminergic neurons² (Fig. 2). The derivation of human ES cells from week-old embryos, containing 100–150 cells, involves the destruction of the embryos. Human ES cell lines can also be diverted to different fates, including nerve and muscle³. So far the embryos from which human ES cells have been derived were almost all donated, either fresh or frozen, by couples undergoing *in vitro* fertilization (IVF) treatment ('supernumerary' or 'spare' embryos). One centre⁴ has derived human ES cell lines from donated oocytes fertilized for the purpose of stem cell derivation. They could also be derived from the transfer of a somatic cell nucleus (somatic cell nuclear transfer or SCNT) into a donated oocyte from which the nuclear material had been removed (Fig. 3). Mouse ES cells have been successfully derived from SCNT embryos⁵. These various sources raise different ethical issues.

Moral value of donated embryos

Couples undergoing IVF treatment often have embryos that they no longer need. They have the option of just letting the embryos die, donating them to another couple, or donating them for research. For those who believe the human embryo from the one-cell stage onwards has absolute moral value, equal to that of a newborn baby or an adult, any embryo research is ethically unacceptable, as it is tantamount to murder. But life is continuous — the egg and the sperm are alive, as well as the one-celled embryo — and although a new genetic constitution comes into being at fertilization, many people feel that moral value develops gradually. The early embryo could be regarded as having symbolic moral value, as a potential human person, and therefore worthy of respect⁶, but in judging whether or not embryo research is ethically acceptable, both the stage of development and the object of the research have to be taken into account. Countries with laws that allow human embryo research (see Tables 1 and 2), set a time limit for research (usually 14 days, just before the fetus begins to form, which is the limit for twinning and hence the start of individual human development). Laws may also specify purposes of clinical relevance. Stem cell



Figure 3 Somatic cell nuclear transfer. An adult cell nucleus is injected (right) into a mouse egg (upper centre) from which the genetic material has been removed. The egg is then stimulated to develop into an embryo from which stem cells can be derived in culture. If human stem cells were derived in this way, using the cell nucleus of a patient, the risk of transplant rejection would be minimized. (Image courtesy of J. King-Holmes/SPL.)

Table 1 Human embryo research (and ES cell derivation) in Europe

Regulated by law	Denmark, Finland, Hungary, Spain, Sweden, United Kingdom
Law in preparation	France*, Netherlands, Portugal
Prohibited	Austria, France*, Germany, Ireland, Norway
No law yet	Belgium, Czech Republic, Greece, Italy, Poland, Slovenia, Switzerland, Turkey

*The 1994 Bioethics Law prohibited human embryo research for a period of 5 years. Prime Minister Lionel Jospin has announced a forthcoming bill which would authorize research on embryos, to include the study of new therapies based on stem cells.

Table 2 Human embryo research (and ES cell derivation) outside Europe

Australia	Victoria prohibits human embryo research and ES cell derivation, but allows research on imported ES cell lines.
Israel	Derivation of human ES cell lines from 'spare' IVF embryos is permitted.
Japan	Human ES cell derivation is to be regulated. Reproductive cloning is prohibited. Other forms of assisted reproductive technology and human embryo research will not be regulated.
United States	A few states have laws prohibiting any human embryo research, otherwise there is no regulation of privately funded research. Congress prohibits federal funding of human embryo research, including ES cell derivation. President Bush has recently said that NIH should fund research on human ES cell lines already in existence, but not on any derived after 9 August 2001.

research, aimed at therapy for serious and intractable diseases, seems eminently relevant.

Because pluripotent cells proliferate indefinitely, in principle it should be possible for much clinically valuable research to be carried out worldwide on the directed differentiation of stem cells, using the relatively few human pluripotent stem cell lines that have been derived so far. This is the basis for the policy proposed by President George W. Bush, who claims that there are already 64 human ES cell lines in existence: the United States National Institutes of Health (NIH) should be allowed to fund research on those, but not on any made subsequent to 9 August 2001. Even if all 64 lines can be used for research, US workers may find this unduly restrictive. Additionally, the extent to which existing stem cell lines have been distributed is limited by issues involving patenting, commercial secrecy and restrictive material transfer agreements. Most countries in Europe are moving towards

allowing derivation of ES cells, as are Japan, Israel, Singapore, India and some other Asian countries. Germany may soon join Victoria (Australia) in allowing research on human ES cell lines that are imported from other countries.

Embryos created for research

Where a research project cannot be carried out on supernumerary embryos, some countries (for example, the United Kingdom) allow the fertilization of donated oocytes for the purpose of the project. Whatever the reason that the embryo was brought into being, it presumably has the same moral value, and is equally destined to perish in a few days whether or not it is used for research. But it is generally agreed that it might be neither necessary nor desirable to make more embryos for stem cell research, if the 'spare' frozen embryos that could be donated for this purpose are sufficient in number and quality.

Research on embryos made by SCNT could not, by definition, be done on 'spare' embryos, whether fresh or frozen. Would this use of SCNT be ethical? The principal objection put forward is that it might facilitate reproductive cloning in humans (that is, cloning people), as reproductive cloning would use similar SCNT technology. But this is not in itself an ethical argument: the ethical objections are to reproductive cloning. The potential benefit of SCNT stem cells is that they could be made from the patient's own somatic cell nuclei, custom-built for immunological compatibility, and so transplant rejection should not occur. The downside is the ethical dilemma presented by the resource issue. Such research would require a supply of unfertilized human eggs, and SCNT is an inefficient process: many eggs would be required to make a single ES cell line. Many IVF centres have long waiting lists of women needing donated eggs to tackle their own infertility problems. In countries where it is permissible to buy eggs (in the United States, an egg donor may be paid thousands of dollars), the question of exploitation of women arises, as egg donation is not without risk. Perhaps use could be made of the many thousands of very immature eggs in every woman's ovary. In animals it has been shown that these can be matured *in vitro* and will support development to the blastocyst stage, allowing the possible derivation of SCNT stem cells.

Clinical perspectives

But even if all problems were overcome, would SCNT stem cells ever be a realistic clinical option? Probably only for the very rich. Any such personalized treatment will always remain labour intensive, and

hence expensive. Other means of circumventing transplant rejection are likely to win out. Immunosuppressive drugs are becoming cheaper and more effective, with fewer side effects. Stem cells could be genetically modified to reduce their foreignness, and methods of inducing specific tolerance in the patient may be developed.

The value of SCNT research may ultimately lie in throwing light on the genetic basis for human diseases, or in enhancing our understanding of reprogramming faulty human genes, rather than making stem cells for transplantation purposes. In addition, if we understood how egg cytoplasm can reprogramme a nucleus, perhaps we could reprogramme the patient's somatic nuclei into stem cells without any need for eggs and embryos. There is also a point of view that SCNT presents no ethical problems because the entity to which it gives rise is not an embryo as defined as the product of fusion between sperm and egg, but an artefact, possessing no moral significance. This is an ethically convenient argument for those countries where all human embryo research is prohibited, but it demands a considerable degree of sophistry. For most people, an embryo is simply the earliest stage of development of an animal or plant. Was Dolly never an embryo? Is Dolly perhaps not a sheep, but merely an ovine cyborg?

Attempts have been made to hype adult stem cells at the expense of ES cells. But almost all scientists working on either adult or ES cells take the view that both lines of research should be energetically pursued. Perhaps some diseases may benefit from one, and some from the other. In addition, in those countries where fertilization of donated eggs is allowed for research purposes, it may be valuable to make ES cell lines free of the genetic bias inherent in left-over IVF embryos. And although ES cells derived from SCNT embryos may never prove economic for routine clinical treatment, they could nonetheless make an immensely valuable contribution to medical research. Let a thousand stem cell lines bloom — but let them bloom in full view of all, so that they can be subject to scientific and ethical review, freely available for research and one day, perhaps, for treating diseases. □

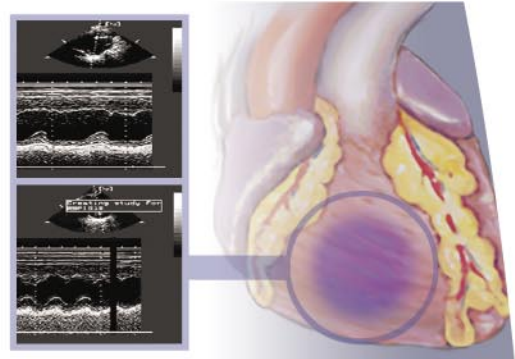
1. Lagasse, E. *et al.* Purified hematopoietic stem cells can differentiate into hepatocytes *in vivo*. *Nature Med.* **6**, 1229–1234 (2000).
2. Lee, S.-H., Lumelsky, N., Studer, L., Auerbach, J. M. & McKay, R. D. Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. *Nature Biotechnol.* **18**, 675–679 (2000).
3. Reubinoff, B. E., Pera, M. F., Fong, C.-Y., Trounson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nature Biotechnol.* **18**, 399–404 (2000).
4. Lanzendorf, S. E. *et al.* Use of human gametes obtained from anonymous donors for the production of human embryonic stem cell lines. *Fert. Steril.* **76**, 132–137 (2001).
5. Munsie, M. J. *et al.* Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. *Curr. Biol.* **10**, 989–992 (2000).
6. Robertson, J. A. Human embryonic stem cell research: ethical and legal issues. *Nature Rev. Genet.* **2**, 74–78 (2001).

Leaders In Stem Cell Medicine

Applications of Mesenchymal Stem Cells

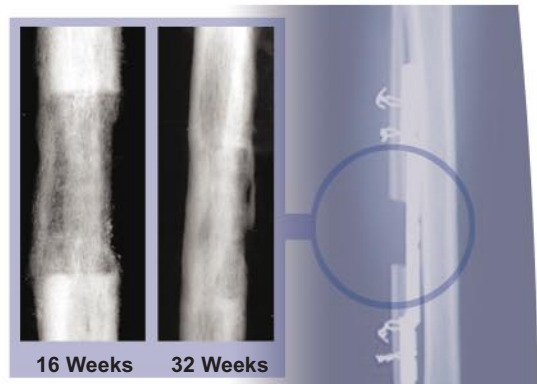
Cardiac Muscle Repair

CardioCel™ is a product candidate that uses "off-the-shelf" MSCs to regenerate damaged or diseased heart muscle. It is targeted primarily to post-myocardial infarct patients. Osiris has demonstrated robust engraftment and cardiac differentiation of allogeneic MSCs in both pig and mouse infarction models. Delivery of allogeneic porcine MSCs improved left ventricular function in a reperfusion infarct model. Clinical trials are anticipated for early 2002.



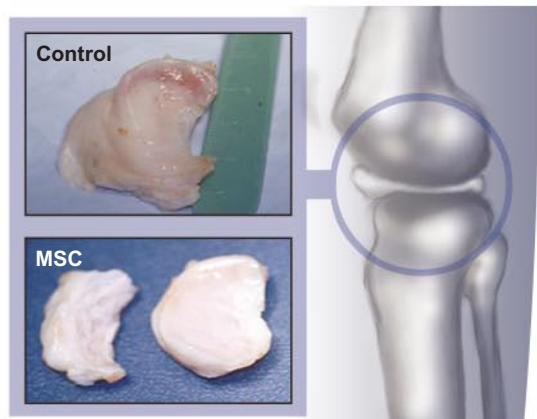
Bone Regeneration

The OsteoCel® product candidate combines off-the-shelf human MSCs with an appropriate biocompatible matrix to regenerate bone that has been injured, diseased or degenerated. The figure illustrates critical gap repair in a baboon model, using allogeneic baboon MSCs on a supporting matrix. A Phase I clinical trial is underway and Phase II trials are planned for early 2002. Osiris seeks to address not only segmental gap cases, but also smaller non-unions, cavitary defects, spinal repair and craniofacial reconstruction.



Joint Repair

Chondrogen™ is a product candidate comprising off-the-shelf human MSCs delivered in liquid suspension by intra-articular injection. The aim is to regenerate a severely damaged meniscus, thus delaying or halting the progression of osteoarthritis in the knee joint. Osiris has established a goat model for the induction of osteoarthritis that includes a complete medial menisectomy. The figure shows the regeneration of the medial meniscus six weeks after a single intra-articular injection of 10 million allogeneic goat MSCs. Clinical trials are anticipated for early 2002.



Osiris develops off-the-shelf cellular products based on the human mesenchymal stem cell (hMSC). These products start with donor material that is expanded in vitro using Osiris' proprietary process. The expanded cells are cryopreserved in the undifferentiated stem cell state, shipped to and stored at the clinical site, then thawed for administration. Once delivered, the cells receive host signals that direct appropriate differentiation. Osiris' basic science and pre-clinical research have confirmed the ability to use MSCs in a fully allogeneic context without MHC matching or immunosuppression. Other clinical applications for hMSCs include support of hematopoietic stem cell transplantation (Phase II clinical trials).



Osiris Therapeutics

For more information, visit us at www.osiristx.com

Copyright © 2001 AE Osiris Therapeutics Incorporated

When The Body Can't Heal Itself

This is an exciting time for the field of stem cell biology. It would have been difficult to predict four years ago the tremendous progress that the field has seen both in basic research discoveries, as well as in therapeutic potential. From the first descriptions of the human embryonic and germ stem cells to the multipotentiality of adult stem cells, dramatic developments have occurred on all fronts. These developments have taken place under heightened awareness from scientists, the medical community, elected officials and interested citizens. All of us know much more about stem cells today than at any previous time, yet the foreseeable future promises even more exciting breakthroughs, including the discovery of new adult stem cells, and greater plasticity of these cells. Osiris is delighted to support this Nature Insight on the biology and potential of stem cells.

Almost 20 years ago the field of stem cell therapeutics was launched with the isolation of the hematopoietic stem cell. It took another 10 years before the mesenchymal stem cell was identified and utilized in the repair of connective tissues in pre-clinical models. Subsequently, fetal neural stem cells were used to treat patients with Parkinson's disease and adult liver stem cells have been identified. The search for other adult stem cells, such as those giving rise to pancreatic islets, continues. Preclinical and clinical studies have shown that the *in vivo* environment provides the necessary cues for tissue-appropriate differentiation and regulation of adult mesenchymal stem cells (MSCs). Additionally, recent studies have shown that MSCs are an "immune privileged" cell type and therefore do not stimulate an immune response *in vitro* or *in vivo* and can be used clinically without immunosuppression. Whether other adult or even embryonic stem cells will demonstrate the same immune privilege remains to be shown. Osiris' scientists and collaborators have demonstrated in several large animal models that the use of fully mismatched, donor-derived MSCs repaired bone, cardiac muscle and meniscus without evidence of immune rejection. The use of stem cells in such an "off-the-shelf" manner could make stem cell based cellular therapies feasible and cost-effective in the acute injury setting as well as in chronic disease states. The era of cellular and tissue regeneration for the treatment of disease and the effects of aging has indeed begun.

Though the articles in this publication can only touch on the possibilities that stem cell biology may provide, we can be certain that continued research will enhance significantly our understanding of human development. Through the dedication of many individuals, research teams and institutions, physicians and healthcare professionals, and in cooperation with industry and regulatory agencies, we are confident that stem cell biology will beneficially impact healthcare for all mankind.

Mark F. Pittenger, Ph.D.
Director, Discovery Research



Osiris Therapeutics

For more information, visit us at www.osiristx.com

Copyright © 2001 AE Osiris Therapeutics Incorporated

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Wells (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Wells (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (6F),

19-1 Haraikatanmachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejpn.com

Japan Manager:

Kate Cowan

naturejobs

A model establishment

Founded in the thirteenth century, the University of Cambridge is one of the world's oldest universities. Over the years it has established an international reputation for academic excellence and innovation. It is fitting, then, that the modernist structures that make up Cambridge Science Park were planned within the Tudor spires of Trinity College.

Cambridge Science Park can trace its origins back to the Labour Party's winning election campaign of 1964, during which party leader Harold Wilson referred to "the white heat of technology". This phrase subsequently came to characterize Wilson's time as prime minister, as he urged universities to forge stronger links with industry in order to receive a return on the government's investment in basic research.

In Cambridge, a university committee elected at that time to heed this advice, and Trinity College — founded by Henry VIII in 1546 — suggested that farmland in its possession would make an ideal home for high-tech companies.

But success for this enterprise was slow in coming. Although it is now widely viewed in Europe as a model for science parks, Trinity's plot of land was slow to attract tenants when it opened in 1973. Eventually, early occupants of the park, such as Cambridge Consultants, spawned numerous smaller companies. In the process, the park's evolution reflected the major technological trends of the time. It started as a home for instrumentation companies, became dominated by information technology and telecommunications companies, and is now more biotech-orientated.

It is possible that both university and science park have endured for so long because they have ensured that they have the capacity to change — a quality that scientists who wish to remain relevant in an ever-shifting job market would do well to cultivate.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Cambridge: a centre of
successful enterprise

p4

WWW.NATUREJOBS.COM

Career centre
Information on the scientific
job market

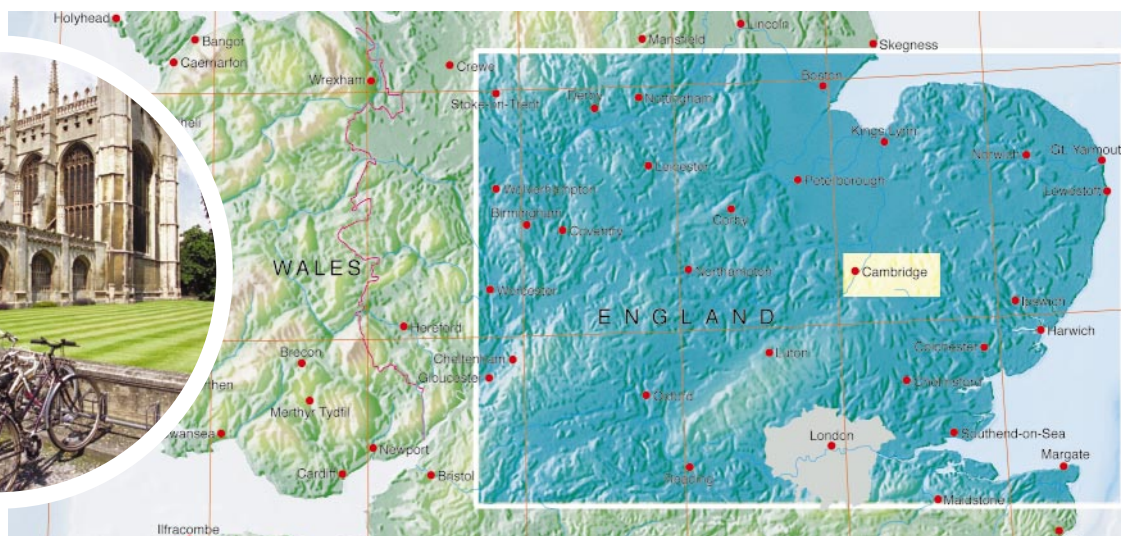
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Managing mergers Cambridge

Job security. Although that phrase no longer means much in the increasingly volatile pharmaceutical sector, it resonates with scientists working in Cambridge — one of the most redundancy-proof regions of Europe for life scientists working in industry.

In the past few years, several pharmaceutical companies have either consolidated their research operation in southeast England following mergers, or shut down facilities entirely. But through all the turbulence, the Cambridge area has maintained

relatively low unemployment compared with the national average — at least in part because employers in the many local science parks have absorbed staff let go by the drugs companies.

Changes in the Cambridge area in recent years have included Roche closing a facility in Welwyn and SmithKline Beecham downsizing a facility in Harlow after merging with Glaxo Wellcome which, in turn, cut its workforce in Stevenage. Area branches of Pfizer and Warner Lambert also reduced the number of employees when they combined.

As drugs companies

consolidate, the number and size of biotech companies in the region has continued to increase. The critical mass created by these small companies, along with larger, more established firms and public-research institutions, means a positive climate for job seekers, a massive talent pool for recruiters and an ever-changing environment for both.

The existence of more than 160 biotech companies in the Cambridge area — many in the area's five research parks — makes employees feel more secure about their future. "They say, 'If it doesn't work out, there are 10 other companies I can join,'" says Jeremy Fairbrother, senior bursar of Cambridge University's Trinity College and director of the Cambridge Science Park.

A CHANGE OF SCENE

The concept of job security has changed greatly in the past few years. Not long ago, people working in the drugs industry expected to stay with one firm for their entire career. Allan Marchington — who is now general manager, Europe, at Millennium Pharmaceuticals in Cambridge, Massachusetts — recalls colleagues' reactions some years ago when he decided to leave a large, established firm to help start Cambridge Combinatorial, a small biotech company. People asked him: "What are you doing leaving? You have a secure job with Pfizer."

Of course, his stint in biotech hasn't exactly been static. Cambridge Combinatorial eventually became Cambridge Drug Discovery, before being bought by Millennium Pharmaceuticals. Millennium's UK branch will shortly complete an addition to its workspace, with plans to construct an even bigger

Mixed fortunes for postdocs

Cambridge postdocs are faced with the classic question of whether to stay in the region where they were trained or to go and seek their fortune elsewhere. How they answer is affected by financial considerations.

Robert Felix, a postdoc in Cambridge University's department of pharmacology, wants to stay. He knows and likes the area, and is curious to explore the biotech climate there. He

is the only one of eight or so academic colleagues to consider sticking around.

He feels industry is his only option, because he had to take on debt to finance his graduate education. Although he is comfortable with Cambridge's biotech scene, he recognizes that he may not be able to participate in it and live comfortably while paying off his loans, as the cost of living in the area is rising sharply. P.S.

building by 2003 — a typical plan in Cambridge, where it seems that every company has plans to move, build or both.

Marchington notes that two general types of people are in demand by the region's biotech companies: senior people from pharmaceuticals who are "disenfranchised by megamergers" and young people with creativity, energy and enthusiasm.

The veterans relish the prospect of a new challenge and the promise of less bureaucracy. Those qualities attracted two principals of RiboTargets — the first company to focus exclusively on developing drugs that target RNA — away from two separate companies that have since combined.

Research director Harry Finch sees his position at RiboTargets as a fitting culmination to a long career in pharmaceuticals. He spent 25 years with Glaxo Wellcome, rising to be director of chemistry before joining RiboTargets in 2001. "I'd always had it in my mind to come into a small company and have some fun," Finch says. "You can be nimble."

But RiboTargets' chief scientific officer, David Knowles, who came from SmithKline Beecham in 1999, notes that applicants for jobs at the new company also ask about its long-term financial security.

STABLE OUTLOOK

In light of all the consolidation, stability is increasingly important to young scientists. Sanger Centre postdoc Charles Cox felt so positive about his Cambridge prospects during a previous postdoctoral stint that he bought a house before he secured his existing post — a job that could easily last another five years. When that appointment runs out, he is confident he can find another without moving house.

Of course, postdocs here, as in other parts of Europe, face additional challenges (see 'Mixed fortunes for postdocs', opposite). And Cox may have been prescient in his purchase — the rising cost of housing is one infrastructure issue that may challenge the long-term vitality of the area (see 'Infrastructure issues loom', right).

One of the traditional advantages of having a company in Cambridge — besides a postcode that offers cachet to venture capitalists — is the access to the scientific ideas that are generated nearby. Sometimes those, like employees, are liberated by mergers.

For example, Trevor Jarman, business development director of Alizyme — a company working on treatments for obesity and gastrointestinal disorders — recalls that SmithKline Beecham shelved renzapride, a potential treatment for irritable bowel syndrome, for strategic reasons. Once the company combined with Glaxo Wellcome, Alizyme snapped up the rights, and the compound is now midway through clinical trials.

The ripples of mergers can even reach academia. David James, chair of Cambridge University's department of pharmacology, notes that the merger of Glaxo Wellcome and SmithKline Beecham will mean that a departmental research unit once funded by Glaxo Wellcome will soon be closed down. The new company no longer feels that funding the programme, which supports as many as three graduate students, is in its interests, says James. But he says there is some optimism that another drugs or a biotech company will step in and support it.

Infrastructure issues loom

Two people whose jobs include promoting Cambridge biotech note that much needs to be done to protect the area. As science parks and employees continue to amass, they bring congestion and unaffordable housing, say Mark Treherne, chair of the Eastern Region Biotechnology Initiative, and Brian Moon, chief executive at Cambridge Consultants.

Treherne notes that one of the area's biggest advantages — easy access to Stansted Airport and London — is being threatened by its success. Although there are good rail links, the area's dependence on cars continues to grow — and the roads weren't built for the present amount of traffic. Journeys that should

take an hour or less are increasingly taking longer. He's not sure that building more roads is the answer, as this might invite more sprawl. "We don't want to turn into central Boston," he says.

Moon notes that rising house prices may keep out some talent — or force people to commute farther, adding to the congestion and detracting from the quality of life that brings many to Cambridge in the first place. The area must recognize that, unchecked, it will end up with the sort of congestion and excessive house prices now seen in San Francisco, he says. "What we seem to be good at is building new science parks quickly. We're not so good at building new housing or better roads."

P.S.

Despite all the volatility, many biotech companies in the area say that the net result of corporate transactions still leaves them short of qualified scientists. Martin Davies, managing director of BioRobotics, which designs and makes automation for molecular biology research, notes that the company can't wait around for drugs companies to merge or other small companies to founder, thus freeing up their labour pool. "Increasingly, we're looking outside the region," he says.

Paul Smaglik is *Naturejobs* editor.



Star turns: Cambridge's science parks have proved to be a great success for the area.